PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/82

A1

(11) International Publication Number: WO 97/12047

(43) International Publication Date: 3 April 1997 (03.04.97)

(21) International Application Number: PCT/US96/16078

(22) International Filing Date: 30 September 1996 (30.09.96)

(30) Priority Data:
08/537,083 29 September 1995 (29.09.95) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/537,083 (CIP)
Filed on 29 September 1995 (29.09.95)

(71) Applicant *(for all designated States except US)*: CALGENE, INC. [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KRIDL, Jean [US/US]; 538 Reed Drive, Davis, CA 95616 (US). KNAUF, Vic, C. [US/US]; 1013 Hillview Lane, Winters, CA 95694 (US). LASSNER, Michael, W. [US/US]; 721 Falcon Avenue, Davis, CA 95616 (US).

(74) Agents: SCHERER, Donna, E. et al.; Calgene, Inc., 1920 Fifth Street, Davis, CA 95616 (US).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PLANT STEAROYL-ACP THIOESTERASE SEQUENCES AND METHODS TO INCREASE STEARATE CONTENT IN PLANT SEED OILS

(57) Abstract

This invention relates to plant thioesterases, specifically plant acyl-ACP thioesterases having significant activity on stearoyl-ACP. DNA constructs useful for the expression of a plant stearoyl-ACP thioesterase in a plant seed cell are described. Such constructs will contain a DNA sequence encoding the plant stearoyl-ACP thioesterase of interest under the control of regulatory elements capable of preferentially directing the expression of the plant stearoyl-ACP thioesterase in seed tissue, as compared with other plant tissues, when such a construct is expressed in a transgenic plant. This invention also relates to methods of using a DNA sequence encoding a plant stearoyl-ACP thioesterase for the modification of the stearate produced in a plant seed cell. A plant thioesterase having significant activity on stearoyl-ACP exemplified herein is from mangosteen. Modified stearate concentration may be obtained by expression of mangosteen thioesterase alone or in combination with antisense constructs for reduction of native stearoyl-ACP desaturase in plant seeds.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan

PLANT STEAROYL-ACP THIOESTERASE SEQUENCES AND METHODS TO INCREASE STEARATE CONTENT IN PLANT SEED OILS

INTRODUCTION

Field of the Invention

This invention relates to the application of genetic engineering techniques to plants. More specifically, the invention relates to plant acyl-ACP thioesterase sequences having substantial activity on C18:0-ACP and methods for the use of such sequences to increase 18:0 levels in plant seed oils.

BACKGROUND

Fatty acids are organic acids having a hydrocarbon chain of from about 4 to 24 carbons. Many different kinds of fatty acids are known which differ from each other in chain length, and in the presence, number and position of double bonds. In cells, fatty acids typically exist in covalently bound forms, the carboxyl portion being referred to as a fatty acyl group. The chain length and degree of saturation of these molecules is often depicted by the formula CX:Y, where "X" indicates number of carbons and "Y" indicates number of double bands.

Fatty acyl groups are major components of many lipids, and their long, non-polar hydrocarbon chain is responsible for the water-insoluble nature of these lipid molecules. The type of covalent linkage of the fatty acyl group to other factors can vary. For example, in biosynthetic reactions they may be covalently bound via a thioester linkage to an acyl carrier protein (ACP) or to CoenzymeA (CoA), depending on the particular enzymatic reaction. In waxes, fatty acyl groups are linked to fatty alcohols via an ester linkage, and triacylglycerols have three fatty acyl groups linked to a glycerol molecule via an ester linkage.

The production of fatty acids in plants begins in the plastid with the reaction between acetyl-CoA and malonyl-ACP to produce butyryl-ACP catalyzed by the enzyme, ß-ketoacyl-ACP

synthase III. Elongation of acetyl-ACP to 16- and 18- carbon fatty acids involves the cyclical action of the following sequence of reactions: condensation with a two-carbon unit from malonyl-ACP to form a ß-ketoacyl-ACP (ß-ketoacyl-ACP synthase), reduction of the keto-function to an alcohol (ß-ketoacyl-ACP reductase), dehydration to form an enoyl-ACP (ß-hydroxyacyl-ACP dehydrase), and finally reduction of the enoyl-ACP to form the elongated saturated acyl-ACP (enoyl-ACP reductase). ß-ketoacyl-ACP synthase I, catalyzes elongation up to palmitoyl-ACP (C16:0), whereas ß-ketoacyl-ACP synthase II catalyzes the final elongation to stearoyl-ACP (C18:0). The longest chain fatty acids produced by the FAS are typically 18 carbons long. A further fatty acid biochemical step occurring in the plastid is the desaturation of stearoyl-ACP (C18:0) to form oleoyl-ACP (C18:1) in a reaction catalyzed by a Δ-9 desaturase, also often referred to as a "stearoyl-ACP desaturase" because of its high activity toward stearate the 18 carbon acyl-ACP.

Carbon-chain elongation in the plastids can be terminated by transfer of the acyl group to glycerol 3-phosphate, with the resulting glycerolipid retained in the plastidial, "prokaryotic", lipid biosynthesis pathway. Alternatively, specific thioesterases can intercept the prokaryotic pathway by hydrolyzing the newly produced acyl-ACPs into free fatty acids and ACP.

Subsequently, the free fatty acids are converted to fatty acyl-CoA's in the plastid envelope and exported to the cytoplasm. There, they are incorporated into the "eukaryotic" lipid biosynthesis pathway in the endoplasmic reticulum which is responsible for the formation of phospholipids, triglycerides and other neutral lipids. Following transport of fatty acyl CoA's to the endoplasmic reticulum, subsequent sequential steps for triglyceride production can occur. For example, polyunsaturated fatty acyl groups such as linoleoyl and ∝-linolenoyl, are produced as the result of sequential desaturation of oleoyl acyl groups by the action of membrane-bound enzymes. Triglycerides are formed by action of the 1-, 2-, and 3- acyl-ACP transferase enzymes glycerol-3-phosphate acyltransferase, lysophosphatidic acid acyltransferase and diacylglycerol acyltransferase. The fatty acid composition of

a plant cell is a reflection of the free fatty acid pool and the fatty acids (fatty acyl groups) incorporated into triglycerides as a result of the acyltransferase activities.

The fatty acid composition of an oil determines its physical and chemical properties, and thus its uses. Plants, especially plant species which synthesize large amounts of oils in plant seeds, are an important source of oils both for edible and industrial uses. Various combinations of fatty acids in the different positions in the triglyceride will alter the properties of triglyceride. For example, if the fatty acyl groups are mostly saturated fatty acids, then the triglyceride will be solid at room temperature. In general, however, vegetable oils tend to be mixtures of different triglycerides. The triglyceride oil properties are therefore a result of the combination of triglycerides which make up the oil, which are in turn influenced by their respective fatty acid compositions.

For example, cocoa-butter has certain desirable qualities (mouth feel, sharp melting point, etc.) which are a function of its triglyceride composition. Cocoa-butter contains approximately 24.4% palmitate (16:0), 34.5% stearate (18:0), 39.1% oleate (18:1) and 2% linoleate (18:2). Thus, in cocoa butter, palmitate-oleate-stearate (POS) comprises almost 50% of triglyceride composition, with stearate-oleate-stearate (SOS) and palmitate-oleate-palmitate (POP) comprising the major portion of the balance at 39% and 16%, respectively, of the triglyceride composition. Other novel oils compositions of interest might include trierucin (three erucic) or a triglyceride with medium chain fatty acids in each position of the triglyceride molecule.

Of particular interest also are triglyceride molecules in which stearate is esterified at the sn-1 and sn-3 positions of a triglyceride molecule and oleate is at sn-2. Vegetable oils rich in such SOS (Stearate-Oleate-Stearate) molecules share certain desirable qualities with cocoa butter yet have a degree of additional hardness when blended with chocolate-based fats. SOS - containing vegetable oils are currently extracted from relatively expensive oilseeds from certain trees grown in tropical areas such as Sal, Shea, and Illipe

trees from India, Africa, and Indonesia respectively. Cheaper and more conveniently grown sources for SOS -type vegetable oils are desirable.

In addition, vegetable oils rich in stearate fatty acid content tend to be solid at room temperature. Such vegetable fats can be used directly in shortenings, margarine and other food "spread" products, obviating the need for chemical hydrogenation. Hydrogenation is a process to convert unsaturated fatty acids in liquid oils to a saturated form which in turn converts the oil into a solid fat useful in margarine and shortening applications. The cost and any other factors associated with chemical hydrogenation, such as the production of trans fatty acids, can be avoided if the vegetable oil is engineered to be stearate rich in the plant seed.

Moreover, some plant tissues use 18 carbon fatty acids as precursors to make other compounds. These include saturated long chain fatty acids longer than 18 carbons in length. Since very little stearate typically accumulates in common oilseed crops, it may be necessary to increase stearate accumulation if one wants to increase production of compounds which depend upon supply of stearate fatty acids for synthesis.

Thus, a variety of plant oils modifications are desired, including alternative crop sources for certain oils products and/or means to provide novel fatty acid compositions for plant seed.

DESCRIPTION OF THE FIGURES

Figure 1. An amino acid sequence alignment of representative Class I (FatA) and Class II (FatB) thioesterases is provided. UcFatB1 (SEQ ID NO:1) is a California bay C12 thioesterase. CcFatB1 (SEQ ID NO:2) is a camphor C14 thioesterase. CpFatB1 (SEQ ID NO:3) is a *Cuphea palustris* C8 and C10 thioesterase. CpFatB2 (SEQ ID NO:4) is a *Cuphea palustris* C14 thioesterase. GarmFatA1 (SEQ ID NO:5) is a mangosteen 18:1 thioesterase which also has considerable activity on C18:0 acyl-ACP substrates. BrFatA1 (SEQ ID NO:6) is an 18:1 thioesterase from *Brassica rapa* (aka *Brassica*

campestris). Amino acid sequences which are identical in all of the represented thioesterases are indicated by bold shading.

Figure 2. The nucleic acid sequence and translated amino acid sequence of a mangosteen Class I acyl-ACP thioesterase cDNA clone, GARM FatA1 (SEQ ID NO:7), are provided.

Figure 3. The nucleic acid sequence and translated amino acid sequence of a mangosteen Class I acyl-ACP thioesterase cDNA clone, GARM FatA2 (SEQ ID NO:8), are provided.

Figure 4. Data from fatty acid composition (weight percent) analyses of T2 mature pooled seed from *Brassica* plants transformed with napin/mangosteen TE construct pCGN5266 are provided.

Figure 5. Fatty acid composition data (weight percent) from analyses of T2 half seeds from events 5266-LP004-2 and 5266-SP30021-29 are provided.

Figure 6. The nucleic acid sequence and translated amino acid sequence of a *Brassica napus* BND11 stearoyl-ACP desaturase cDNA clone are provided.

Figure 7. The nucleic acid sequence and translated amino acid sequence of a *Brassica napus* BND9 stearoyl-ACP desaturase cDNA clone are provided.

SUMMARY OF THE INVENTION

This invention relates to plant thioesterases, specifically plant acyl-ACP thioesterases having substantial activity on 18:0-ACP substrates such that the C18:0 content in a target plant seed oil can be dramatically increased upon expression of the plant acyl-ACP thioesterase in the seeds of the target plant.

DNA constructs useful for the expression in a plant seed cell of a plant acyl-ACP thioesterase having substantial activity on 18:0-ACP substrates are described herein. Such constructs will contain a DNA sequence encoding the plant acyl-ACP thioesterase under the control of regulatory elements capable of preferentially directing the expression of the plant acyl-ACP thioesterase in seed tissue, as compared with other plant tissues. At least one element of the DNA

construct will be heterologous to at least one other of the elements, or when found in a plant cell, at least one element will be heterologous to the plant cell. Of particular interest for use in the constructs of the present invention is a Class I acyl-ACP thioesterase, Garm FatA1, obtained from mangosteen (*Garcinia mangifera*).

In yet a different embodiment, host plant cells containing a first DNA construct capable of expressing a plant acyl-ACP thioesterase having substantial activity on 18:0-ACP substrates and a second DNA construct capable of expressing an anti-sense stearoyl-acyl ACP desaturase sequence are desired. Such a first DNA construct will contain a DNA sequence encoding the plant acyl-ACP thioesterase of interest under the control of regulatory elements capable of preferentially directing the expression of the plant acyl-ACP thioesterase in seed tissue as compared with other plant tissues when such a construct is expressed in a transgenic plant. The second DNA construct will contain a DNA sequence encoding a plant stearoyl-acyl ACP desaturase element positioned in an anti-sense orientation under the control of regulatory elements capable of directing the transcription of the plant stearoyl-acyl ACP desaturase in the plant host cell.

In a different aspect, this invention relates to methods of using a DNA sequence encoding a plant acyl-ACP thioesterase for modifying the composition of triglycerides, i.e., plant oils, produced by a plant seed. Such modified oil compositions may be obtained either by expression of the acyl-ACP thioesterase alone, or by expression of the acyl-ACP thioesterase in combination with a second construct which provides for reduction of the level of the native stearoyl-ACP desaturase of the target plant species. Plants and plant parts, particularly seeds and oils extracted from such seeds, having such a modified fatty acid composition are contemplated herein.

Also provided in the present invention is a novel antisense desaturase plant expression construct which provides regions from two different classes of desaturase genes from a multigene family. Expression of this construct in transgenic plant seeds provides an improved method for increasing

stearate levels in plant seeds by antisense methods. Higher levels of stearate are obtained as a result of this improved method, as well as higher percentages of first generation transformants which display the desired increased stearate phenotype.

DETAILED DESCRIPTION OF THE INVENTION

By this invention, a mechanism for the increased accumulation of stearate (C18:0) in plants is provided. As described herein, plant acyl-ACP thioesterases having substantial activity toward 18:0-ACP substrates are involved in the accumulation of stearate in at least some plant species and may be expressed in transgenic plant seeds to provide an increase in the levels of stearate fatty acids. Furthermore, it is demonstrated that such 18:0-ACP thioesterases are members of the Class I group of plant acyl-ACP thioesterases.

Class I type thioesterases have been found in essentially all plant sources examined to date, and are suggested to be essential "housekeeping" enzymes (Jones *et al. The Plant Cell* (1995) 7:359-371) required for membrane biosynthesis. Class I type thioesterases have previously been shown to have activity primarily on 18:1 acyl-ACP substrates, with some lesser amount of activity on 16:0 substrates, and only little or no activity on 18:0 substrates. Examples of Class I thioesterases from safflower, *Cuphea hookeriana* and *Brassica rapa* (*campestris*), which have activity primarily on 18:1-ACP substrate, have been described (WO 92/20236 and WO 94/10288). Other 18:1 thioesterases have been reported in *Arabidopsis thaliana* (Dormann *et al.* (1995) *Arch. Biochem. Biophys. 316*:612-618), *Brassica napus* (Loader *et al.* (1993) *Plant Mol. Biol. 23*:769-778) and coriander (Dormann *et al.* (1994) *Biochem. Biophys. Acta 1212*:134-136). A Class I thioesterase from soybean (WO 92/11373) was reported to provide 10- and 96-fold increases in 16:0-ACP and 18:1-ACP activity upon expression in *E. coli*, and a smaller (3-4 fold) increase in 18:0-ACP activity.

A second class of plant thioesterases, Class II (or FatB) thioesterases, include enzymes that primarily utilize fatty acids with shorter chain-lengths, from C8:0 to C14:0 (medium chain fatty acids) as well as C16:0. Class II thioesterases

preferably catalyze the hydrolysis of substrates containing saturated fatty acids. Class II (or FatB) thioesterase genes have been isolated from California Bay, elm, *Cuphea hookeriana*, *Cuphea palustris*, *Cuphea lanceolata*, nutmeg, *Arabidopsis thaliana*, mango, leek and camphor. A FatB thioesterase gene was also identified in mangosteen in the gene isolation experiments described herein. Expression of the FatB gene in *E. coli* demonstrated hydrolysis activity primarily on 16:0 substrates.

In the following examples, isolation of genes encoding Class I acyl-ACP thioesterases from mangosteen is described. Two different types of Class I thioesterase genes were discovered. One mangosteen thioesterase gene, (GarmFatA2), is shown herein to be an 18:1-ACP specific Class I thioesterase similar to those discovered previously in other plant tissues. However, a second type of mangosteen Class I thioesterase gene, represented by clone GarmFatA1, was discovered which demonstrates 18:1-ACP thioesterase activity (100-fold increase upon expression in *E. coli*), but also demonstrates substantial activity on 18:0-ACP substrates. The 18:0 activity of GarmFatA1 is approximately 25% of the 18:1 activity, whereas in most Class I thioesterases analyzed to date, the 18:1 activity is highly predominant, with activity on 16:0 and 18:0 substrates detectable at less than 5% of the 18:1 activity levels. Additionally, most plant Class I thioesterases demonstrate approximately equal activity on 16:0 and 18:0 substrates, whereas the GarmFatA1 mangosteen thioesterase of the present invention demonstrates a clear preference for hydrolysis of 18:0 substrates over 16:0 substrates.

The novel mangosteen GarmFatA1 thioesterase clone will thus be useful for production of the 18:1/18:0 thioesterase in host cells, and particularly for expression in plant seed cells for modification of TAG fatty acid composition to provide increased levels of C18:0 fatty acyl groups. Furthermore, the mangosteen GarmFatA1 clone will be useful in plant genetic engineering application in conjunction with plants containing elevated levels of C18:0 (stearate) fatty acids. Such plants may be obtained by antisense gene regulation of stearoyl-ACP desaturase in *Brassica* seeds as

described by Knutzon *et al.* (*Proc. Nat. Acad. Sci.* (1992) *89*:2624-2628), and may also be obtained by co-suppression using sense expression constructs of the stearoyl-ACP desaturase gene, or by conventional mutation and plant breeding programs.

The determination that Class I type plant thioesterases are active in the *in vivo* accumulation of 18:0 fatty acids suggests several possibilities for additional plant sources of genes which encode thioesterase proteins having substantial activity on 18:0 substrates. Stearate is found in some natural plant species, particularly tropical plant species, in abundance. For example, other species in the genus *Garcinia* accumulate triglycerides containing stearate in their seeds, e.g., kokum. Other natural plant source of C18:0 fatty acids include plants of the *Mangifera* family: e.g., mango, *Butyrospermum* (shea), *Pentadesma* (tallow tree), *Illipe* (illipe butter), *Theobroma* (cocoa), *Simarouba* (tree of paradise) and *Shorea* (sal).

A plant acyl-ACP thioesterase DNA sequence useful for alteration of stearate levels as described herein encodes for amino acids, in the form of a protein, polypeptide or peptide fragment, which amino acids demonstrate substantial activity on 18:0 acyl-ACP substrates to form 18:0 free fatty acid (i.e., stearate) under plant enzyme reactive conditions. By "enzyme reactive conditions" is meant that any necessary conditions are available in an environment (i.e., such factors as temperature, pH, lack of inhibiting substances) which will permit the enzyme to function.

One skilled in the art will readily recognize that antibody preparations, nucleic acid probes (DNA and RNA) and the like may be prepared and used to screen and recover "homologous" or "related" Class I acyl-ACP thioesterases from a variety of plant sources. Typically, nucleic acid probes are labeled to allow detection, preferably with radioactivity although enzymes or other methods may also be used. As plant thioesterase genes are known to contain extensive sequence homology, various DNA screening methods such as PCR or DNA hybridization methods may be used to identify related Class I thioesterases. Plant thioesterase genes show at least

approximately 50% sequence identity at the nucleic acid level. Between members of Class I thioesterases the percentage of sequence identity is as high as 70-80%, and Class II thioesterases typically demonstrate sequence identity of at least 60%.

Thus, in order to obtain additional stearoyl-ACP thioesterases, a genomic or other appropriate library prepared from the candidate plant source of interest is probed with conserved sequences from one or more Class I plant thioesterases to identify homologously related sequences. Positive clones are then analyzed by restriction enzyme digestion and/or sequencing. When a genomic library is used, one or more sequences may be identified providing both the coding region, as well as the transcriptional regulatory elements of the thioesterase gene from such plant source. Probes can also be considerably shorter than the entire sequence. Oligonucletides may be used, for example, but should be at least about 10, preferably at least about 15, more preferably at least 20 nucleotides in length. When shorter length regions are used for comparison, a higher degree of sequence identity is required than for longer sequences. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified. (*See*, Gould, *et al.*, *PNAS USA* (1989) *86*:1934-1938.)

When longer nucleic acid fragments are employed (>100 bp) as probes, especially when using complete or large cDNA sequences, one can still screen with moderately high stringencies (for example using 50% formamide at 37°C with minimal washing) in order to obtain signal from the target sample with 20-50% deviation, i.e., homologous sequences. (For additional information regarding screening techniques see Beltz, *et al. Methods in Enzymology* (1983) *100*:266-285.).

Once the related acyl-ACP thioesterase sequence is obtained, expression of the plant acyl-ACP thioesterase in a host cell may be obtained to further characterize the activity of the thioesterase. In this manner, additional plant acyl-ACP thioesterases having substantial activity on stearoyl-ACP may be identified for use in the methods of the present

invention. As demonstrated herein by increasing the amount of stearoyl-ACP preferring thioesterase available to the plant FAS complex, an increased percentage of stearate may be provided. Additionally, by decreasing the amount of stearoyl-ACP desaturase available to the plant FAS complex in conjunction with an increase of the amount of stearoyl-ACP thioesterase available, a more marked increased percentage of stearate may be obtained.

The nucleic acid sequences which encode plant stearoyl-ACP thioesterases may be used in various constructs, for example, as probes to obtain further sequences. Alternatively, these sequences may be used in conjunction with appropriate regulatory sequences to increase levels of the respective thioesterase of interest in a host cell for recovery or study of the enzyme *in vitro* or *in vivo* or to decrease levels of the respective thioesterase of interest for some applications when the host cell is a plant entity, including plant cells, plant parts (including but not limited to seeds, cuttings or tissues) and plants.

A nucleic acid sequence encoding a plant acyl-ACP thioesterase of this invention which has substantial activity on 18:0 acyl groups may include genomic, cDNA or mRNA sequence. A cDNA sequence may or may not contain pre-processing sequences, such as transit peptide sequences. Transit peptide sequences facilitate the delivery of the protein to a given organelle and are cleaved from the amino acid moiety upon entry into the organelle, releasing the "mature" sequence. The use of the precursor plant acyl-ACP thioesterase DNA sequence containing the transit peptide and mature protein encoding sequences is preferred in plant cell expression cassettes. Other plastid transit peptide sequences, such as a transit peptide of seed ACP, may be employed to translocate the plant stearoyl-ACP thioesterase of this invention to various organelles of interest.

Furthermore, as discussed above the complete genomic sequence of the plant stearoyl-ACP thioesterase may be obtained by the screening of a genomic library with a probe, such as a cDNA probe, and isolating those sequences which regulate expression in seed tissue. In this manner, the

transcription and translation initiation regions, introns, and/or transcript termination regions of the plant stearoyl-ACP thioesterase may be obtained for use in a variety of DNA constructs, with or without the thioesterase structural gene.

Once the desired plant stearoyl-ACP thioesterase nucleic acid sequence is obtained, it may be manipulated in a variety of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, transversions, deletions, and insertions may be performed on the naturally occurring sequence. In addition, all or part of the sequence may be synthesized. In the structural gene, one or more codons may be modified to provide for a modified amino acid sequence, or one or more codon mutations may be introduced to provide for a convenient restriction site or other purpose involved with construction or expression. The structural gene may be further modified by employing synthetic adapters, linkers to introduce one or more convenient restriction sites, or the like.

The nucleic acid or amino acid sequences encoding a plant stearoyl-ACP thioesterase of this invention may be combined with other non-native, or "heterologous", sequences in a variety of ways. By "heterologous" sequences is meant any sequence which is not naturally found joined to the plant stearoyl-ACP thioesterase, including, for example, combinations of nucleic acid sequences from the same plant which are not naturally found joined together.

The DNA sequence encoding a plant stearoyl-ACP thioesterase of this invention may be employed in conjunction with all or part of the gene sequences normally associated with the thioesterase. In its component parts, a DNA thioesterase encoding sequence is combined in a DNA construct having, in the 5' to 3' direction of transcription, a transcription initiation control region capable of promoting transcription and translation in a host cell, the DNA sequence encoding plant stearoyl-ACP thioesterase and a transcription and translation termination region.

Potential host cells include both prokaryotic and eukaryotic cells. A host cell may be unicellular or found in

a multicellular differentiated or undifferentiated organism depending upon the intended use. Cells of this invention may be distinguished by having a plant stearoyl-ACP thioesterase foreign to the wild-type cell present therein, for example, by having a recombinant nucleic acid construct encoding a plant stearoyl-ACP thioesterase therein.

Depending upon the host, the regulatory regions will vary, including regions from viral, plasmid or chromosomal genes, or the like. For expression in prokaryotic or eukaryotic microorganisms, particularly unicellular hosts, a wide variety of constitutive or regulatable promoters may be employed. Expression in a microorganism can provide a ready source of the plant enzyme.

For the most part, the constructs will involve regulatory regions functional in plants which provide for modified production of plant stearoyl-ACP thioesterase, and modification of the fatty acid composition. The open reading frame, coding for the plant stearoyl-ACP thioesterase or functional fragment thereof will be joined at its 5' end to a transcription initiation regulatory region such as the wild-type sequence naturally found 5' upstream to the thioesterase structural gene, or to a heterologous regulatory region from a gene naturally expressed in plant tissues. Examples of useful plant regulatory gene regions include those from T-DNA genes, such as nopaline or octopine synthase, plant virus genes, such as CaMV 35S, or from native plant genes.

For such applications when 5' upstream non-coding regions are obtained from other genes regulated during seed maturation, those preferentially expressed in plant embryo tissue, such as ACP and napin-derived transcription initiation control regions, are desired. Such "seed-specific promoters" may be obtained and used in accordance with the teachings of USPN 5,420,034 having a title "Seed-Specific Transcriptional Regulation". Transcription initiation regions which are preferentially expressed in seed tissue, i.e., which are undetectable in other plant parts, are considered desirable for fatty acid modifications in order to minimize any disruptive or adverse effects of the gene product.

Regulatory transcript termination regions may be provided in DNA constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the plant stearoyl-ACP thioesterase or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region which is naturally associated with the transcript initiation region. Where the transcript termination region is from a different gene source, it will contain at least about 0.5 kb, preferably about 1-3 kb of sequence 3' to the structural gene from which the termination region is derived.

Plant expression or transcription constructs having a plant stearoyl-ACP thioesterase as the DNA sequence of interest for increased or decreased expression thereof may be employed with a wide variety of plant life, particularly, plant life involved in the production of vegetable oils for edible and industrial uses. Most especially preferred are temperate oilseed crops. Plants of interest include, but are not limited to, rapeseed (Canola and High Erucic Acid varieties), sunflower, safflower, cotton, *Cuphea*, soybean, peanut, coconut and oil palms, and corn. Depending on the method for introducing the recombinant constructs into the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques.

The method of transformation is not critical to the instant invention; various methods of plant transformation are currently available. As newer methods are available to transform crops, they may be directly applied hereunder. For example, many plant species naturally susceptible to *Agrobacterium* infection may be successfully transformed via tripartite or binary vector methods of *Agrobacterium* mediated transformation. In addition, techniques of microinjection, DNA particle bombardment, electroporation have been developed which allow for the transformation of various monocot and dicot plant species.

In developing the DNA construct, the various components of the construct or fragments thereof will normally be

inserted into a convenient cloning vector which is capable of replication in a bacterial host, e.g., *E. coli*. Numerous vectors exist that have been described in the literature. After each cloning, the plasmid may be isolated and subjected to further manipulation, such as restriction, insertion of new fragments, ligation, deletion, insertion, resection, etc., so as to tailor the components of the desired sequence. Once the construct has been completed, it may then be transferred to an appropriate vector for further manipulation in accordance with the manner of transformation of the host cell.

Normally, included with the DNA construct will be a structural gene having the necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or more markers may be employed, where different conditions for selection are used for the different hosts. A number of markers have been developed for use for selection of transformed plant cells, such as those which provide resistance to various antibiotics, herbicides, or the like. The particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

As mentioned above, the manner in which the DNA construct is introduced into the plant host is not critical to this invention. Any method which provides for efficient transformation may be employed. Various methods for plant cell transformation include the use of Ti- or Ri-plasmids, microinjection, electroporation, DNA particle bombardment, liposome fusion, DNA bombardment or the like. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens*

or *A. rhizogenes* as a mode for transformation, although the T-DNA borders may find use with other modes of transformation.

Once a transgenic plant is obtained which is capable of producing seed having a modified fatty acid composition, traditional plant breeding techniques, including methods of mutagensis, may be employed to further manipulate the fatty acid composition. Alternatively, additional foreign fatty acid modifying DNA sequence may be introduced via genetic engineering to further manipulate the fatty acid composition.

One may choose to provide for the transcription or transcription and translation of one or more other sequences of interest in concert with the expression of a plant stearoyl-ACP thioesterase in a plant host cell. In particular, the reduced expression of stearoyl-ACP desaturase in combination with expression of a plant stearoyl-ACP thioesterase may be preferred in some applications.

Stearoyl-ACP thioesterases may also be used in combination with other thioesterase genes with differing specificities. For example, a transgenic oilseed crop expressing both a lauroyl-ACP thioesterase (WO 92/20236) and a stearoyl-ACP thioesterase during seed development will produce an oil enriched in both lauric and stearic acids. Similarly, a transgenic oilseed crop expressing both a palmitoyl-ACP thioesterase (WO 95/13390) and a stearoyl-ACP thioesterase during seed development will produce an oil enriched in both palmitic and stearic acids. These thioesterase gene constructs may be linked to each other in the genome of the transgenic plant or may be unlinked. Conversely the thioesterase genes may be combined in the same transgenic plant by generating plants with one or the other thioesterase gene and subsequently crossing two plants, one of each type. By selecting parents for such crosses, it is possible to further manipulate the relative ratios of desired fatty acids in a seed oil.

When one wishes to provide a plant transformed for the combined effect of more than one nucleic acid sequence of interest, typically a separate nucleic acid construct will be provided for each. The constructs, as described above contain transcriptional or transcriptional and translational

regulatory control regions. The constructs may be introduced into the host cells by the same or different methods, including the introduction of such a trait by crossing transgenic plants via traditional plant breeding methods, so long as the resulting product is a plant having both characteristics integrated into its genome.

By decreasing the amount of stearoyl-ACP desaturase, an increased percentage of saturated fatty acids is provided. Using anti-sense, transwitch, ribozyme or some other stearoyl-ACP desaturase reducing technology, a decrease in the amount of stearoyl-ACP desaturase available to the plant cell is produced, resulting in a higher percentage of saturates such as one or more of laurate (C12:0), myristate (C14:0), palmitate (C16:0), stearate (C18:0), arachidate (C20:0), behenate (C22:0) and lignocerate (C24:0). In rapeseed reduced stearoyl-ACP desaturase results in increased stearate levels and total saturates (Knutzon *et al.* (1992) *Proc. Nat. Acad. Sci 89*:2264-2628). A novel construct is also provided herein which may be used in antisense reduction of stearoyl-ACP desaturase to increase both the percentage of stearate which may be obtained and the percentage of primary transformants expressing the increased stearate trait.

Of special interest is the production of triglycerides having increased levels of stearate or palmitate and stearate. In addition, the production of a variety of ranges of such saturates is desired. Thus, plant cells having lower and higher levels of stearate fatty acids are contemplated. For example, fatty acid compositions, including oils, having a 10% level of stearate as well as compositions designed to have up to an appropriate 60% level of stearate or other such modified fatty acid(s) composition are contemplated.

Oils with increased percentages of stearate, are desired. Increased stearate percentages (by weight) ranging from native levels to increases of up to 25 fold are described. By manipulation of various aspects of the DNA constructs (e.g., choice of promoters, number of copies, etc.) and traditional breeding methods, one skilled in the art may achieve even greater levels of stearate. By combination of the plant stearoyl-ACP desaturase sequence in combination with the

expression of a plant stearoyl-ACP thioesterase in seed tissue, an increased percentage of stearate can be achieved in rapeseed and other plant species. DNA sequence of *C. tinctorius* stearoyl-ACP desaturase gene, as well as DNA sequences of stearoyl-ACP desaturase genes from a *Ricinus*, a *Brassica* and a *Simmondsia* plant are found in WO 91/13972.

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

Example 1 Mangosteen Thioesterase Gene Sequences

A cDNA bank is prepared from seeds extracted from mature mangosteen fruit using the methods as described in Stratagene Zap cDNA synthesis kit (Stratagene; La Jolla, CA). Oil analysis of the mangosteen tissues used for RNA isolation reveals 18:0 levels of approximately 50%. Oil analysis of seeds from less mature mangosteen fruit reveals 18:0 levels of 20-40%. Total RNA is isolated from the mangosteen seeds by modifying the CTAB DNA isolation method of Webb and Knapp (*Plant Mol. Biol. Reporter* (1990) *8*:180-195). Buffers include:

REC: 50 mM TrisCl pH 9, 0.7 M NaCl, 10 mM EDTA pH8, 0.5% CTAB.

REC+: Add B-mercaptoethanol to 1% immediately prior to use.

RECP: 50 mM TrisCl pH9, 10 mM EDTA pH8, and 0.5% CTAB.

RECP+: Add B-mercaptoethanol to 1% immediately prior to use.

For extraction of 1 g of tissue, 10ml of REC+ and 0.5 g of PVPP is added to tissue that has been ground in liquid nitrogen and homogenized. The homogenized material is centrifuged for 10 min at 12000 rpm. The supernatant is poured through miracloth onto 3ml cold chloroform and homogenized

again. After centrifugation, 12,000 RPM for 10 min, the upper phase is taken and its volume determined. An equal volume of RECP+ is added and the mixture is allowed to stand for 20 min. at room temperature. The material is centrifuged for 20 min. at 10,000 rpm twice and the supernatant is discarded after each spin. The pellet is dissolved in 0.4 ml of 1 M NaCl (DEPC) and extracted with an equal volume of phenol/chloroform. Following ethanol precipitation, the pellet is dissolved in 1 ml of DEPC water.

Briefly, the cloning method for cDNA synthesis is as follows. First strand cDNA synthesis is according to Stratagene Instruction Manual with some modifications according to Robinson, *et al.* (*Methods in Molecular and Cellular Biology* (1992) *3*:118-127). In particular, approximately 57µg of LiCl precipitated total RNA was used instead of 5µg of poly(A)+ RNA and the reaction was incubated at 45°C rather than 37°C for 1 hour.

Probes for library screening are prepared by PCR from mangosteen cDNA using oligonucleotides to conserved plant acyl-ACP thioesterase regions. Probe Garm 2 and Garm 106 are prepared using the following oligonucleotides. The nucleotide base codes for the below oligonucleotides are as follows:

A = adenine	C = cytosine
T = thymine	U = uracil
G = guanine	S = guanine or cytosine
K = guanine or thymine	W = adenine or thymine
M = adenine or cytosine	R = adenine or guanine
Y = cytosine or thymine	
B = guanine, cytosine or thymine	
H = adenine, cytosine or thymine	
N = adenine, cytosine, guanine or thymine	

Garm 2

4874: 5' CUACUACUACUASYNTVNGYNATGATGAA 3' (SEQ ID NO:9)

4875: 5' CAUCAUCAUCAURCAYTCNCKNCKRTANTC 3' (SEQ ID NO:10)

Primer 4874 is a sense primer designed to correspond to possible encoding sequences for conserved peptide

V/L/A W/S/Y V/A M M N, where the one letter amino acid code is used and a slash between amino acids indicates more than one amino acid is possible for that position. Primer 4875 is an antisense primer designed to correspond to possible encoding sequences for peptide D/E Y R R E C.

Garm 106

5424: 5' AUGGAGAUCUCUGAWCRBTAYCCTAMHTGGGGWGA 3' (SEQ ID NO:11)

5577: 5' ACGCGUACUAGUTTNKKNCKCCAYTCNGT 3' (SEQ ID NO:12)

Primer 5424 is a sense primer designed to correspond to possible encoding sequences for peptide E/D H/R Y P K/T W G D. Primer 5577 is an antisense primer designed to correspond to possible encoding sequences for peptide T E W R K/P K.

The DNA fragments resulting from the above reactions are amplified for use as probes by cloning or by further PCR and radiolabeled by random or specific priming.

Approximately 800,000 plaques are plated according to manufacturer's directions. For screening, plaque filters are prehybridized at room temperature in 50% formamide, 5X SSC, 10X Denhardt's, 0.1% (w/v) SDS, 5mM Na_2EDTA, 0.1mg/ml denatured salmon sperm DNA. Hybridization with a mixture of the Garm 2 and Garm 106 probes is conducted at room temperature in the same buffer as above with added 10%(w/v) dextran sulfate and probe. Plaque purification and phagemid excision were conducted as described in Stratagene Zap cDNA Synthesis Kit instructions.

Approximately 90 acyl-ACP thioesterase clones were identified and sorted as to thioesterase type by DNA sequencing and/or PCR analysis. Of the analyzed clones, at least 28 were Class I (FatA) types, and 59 were Class II (FatB) types. Two subclasses of FatA type clones were observed, the most prominent type is termed GarmFatA1 and the single clone of the second subclass is termed GarmFatA2. DNA and translated amino acid sequence of GarmFatA1 clone C14-4 (pCGN5252) (SEQ ID NO:7) is presented in Figure 2. DNA sequence and translated amino acid sequence of the FatA2 clone C14-3 (SEQ ID NO:8) is presented in Figure 3.

Constructs for expression of the Figure 2 Garm FatA1 clone in *E. coli* are prepared as follows. Restriction sites

are inserted by PCR mutagenesis at amino acid 49 (*Sac*I), which is near the presumed mature protein amino terminus, and following the stop codon for the protein encoding region (*Bam*HI). The mature protein encoding region is inserted as a *Sac*I/*Bam*HI fragment into pBC SK (Stratagene; La Jolla, CA) resulting in pCGN5247, which may be used to provide for expression of the mangosteen thioesterase as a *lac*Z fusion protein.

Results of thioesterase activity assays on mangosteen Class I thioesterase clone GarmFatA1 using 16:0, 18:0 and 18:1 acyl-ACP substrates are shown below.

Acyl-ACP Thioesterase activity (cpm/min)

	16:0	18:0	18:1
Control	1400	3100	1733
GarmFatA1	4366	23916	87366

The GarmFatA1 cone demonstrates preferential activity on C18:1 acyl-ACP substrate, and also demonstrates substantial activity (approximately 25% of the 18:1 activity) on C18:0 acyl-ACP substrates. Only a small increase in C16:0 activity over activity in control cells is observed, and the 16:0 activity represents only approximately 3% of the 18:1 activity.

Expression of GarmFatA2 thioesterase in *E. coli* and assay of the resultant thioesterase activity demonstrates that C18:1 is highly preferred as the acyl-ACP substrate. The thioesterase activity on 16:0 and 18:0 acyl-ACP substrates are approximately equal and represent less than 5% of the observed 18:1 activity.

Expression of Class II type mangosteen thioesterase clones in *E.coli* demonstrates that 16:0 is highly preferred over other acyl-ACP substrates.

Example 2 Plant Transformation methods

A. *Agrobacterium*-mediated Transformation

Methods which may be used for *Agrobacterium*-mediated transformation of *Brassica* are described by Radke *et al.*

(*Theor. Appl. Genet.* (1988) *75*:685-694; *Plant Cell Reports* (1992) *11*:499-505).

Transgenic *Arabidopsis thaliana* plants may be obtained by *Agrobacterium*-mediated transformation as described by Valverkens *et al.*, (*Proc. Nat. Acad. Sci.* (1988) *85*:5536-5540).

B. Particle Bombardment

DNA sequences of interest may be introduced as expression cassettes, comprising at least a promoter region, a gene of interest, and a termination region, into a plant genome via particle bombardment as described for example in European Patent Application 332 855 and in co-pending application USSN 07/225,332, filed July 27, 1988.

Briefly, tungsten or gold particles of a size ranging from 0.5mM-3mM are coated with DNA of an expression cassette. This DNA may be in the form of an aqueous mixture or a dry DNA/particle precipitate. Tissue used as the target for bombardment may be from cotyledonary explants, shoot meristems, immature leaflets, or anthers.

The bombardment of the tissue with the DNA-coated particles is carried out using a Biolistics‰ particle gun (Dupont; Wilmington, DE). The particles are placed in the barrel at variable distances ranging from 1cm-14cm from the barrel mouth. The tissue to be bombarded is placed beneath the stopping plate; testing is performed on the tissue at distances up to 20cm. At the moment of discharge, the tissue is protected by a nylon net or a combination of nylon nets with mesh ranging from 10mM to 300mM.

Following bombardment, plants may be regenerated following the method of Atreya, *et al.*, (*Plant Science Letters* (1984) *34*:379-383). Briefly, embryo axis tissue or cotyledon segments are placed on MS medium (Murashige and Skoog, *Physio. Plant.* (1962) *15*:473) (MS plus 2.0 mg/l 6-benzyladenine (BA) for the cotyledon segments) and incubated in the dark for 1 week at 25 ± 2°C and are subsequently transferred to continuous cool white fluorescent light (6.8 W/m^2). On the 10th day of culture, the plantlets are transferred to pots containing sterile soil, are kept in the shade for 3-5 days are and finally moved to greenhouse.

The putative transgenic shoots are rooted. Integration of exogenous DNA into the plant genome may be confirmed by various methods know to those skilled in the art.

Example 3 Construct for Expression of Garm FatA1 Thioesterase in Plant Seeds

The napin expression cassette, pCGN1808, is described in USPN 5,420,034 which is incorporated herein by reference. pCGN1808 is modified to contain flanking restriction sites to allow movement of only the expression sequences and not the antibiotic resistance marker to binary vectors such as pCGN1557 (McBride and Summerfelt (1990) *Plant Mol. Biol.* *14*:269-276). Synthetic oligonucleotides containing *Kpn*I, *Not*I and *Hin*dIII restriction sites are annealed and ligated at the unique *Hin*dIII site of pCGN1808, such that only one *Hin*dIII site is recovered. The resulting plasmid, pCGN3200 contains unique *Hin*dIII, *Not*I and *Kpn*I restriction sites at the 3'-end of the napin 3'-regulatory sequences as confirmed by sequence analysis.

The majority of the napin expression cassette is subcloned from pCGN3200 by digestion with *Hin*dIII and *Sac*I and ligation to *Hin*dIII and *Sac*I digested pIC19R (Marsh, *et al.* (1984) *Gene* *32*:481-485) to make pCGN3212. The extreme 5'-sequences of the napin promoter region are reconstructed by PCR using pCGN3200 as a template and two primers flanking the *Sac*I site and the junction of the napin 5'-promoter and the pUC backbone of pCGN3200 from the pCGN1808 construct. The forward primer contains *Cla*I, *Hin*dIII, *Not*I, and *Kpn*I restriction sites as well as nucleotides 408-423 of the napin 5'-sequence (from the EcoRV site) and the reverse primer contains the complement to napin sequences 718-739 which include the unique *Sac*I site in the 5'-promoter. The PCR was performed using in a Perkin Elmer/Cetus thermocycler according to manufacturer's specifications. The PCR fragment is subcloned as a blunt-ended fragment into pUC8 (Vieira and Messing (1982) *Gene* *19*:259-268) digested with *Hinc*II to give pCGN3217. Sequence of pCGN3217 across the napin insert verifies that no improper nucleotides were introduced by PCR. The napin 5-sequences in pCGN3217 are ligated to the remainder

of the napin expression cassette by digestion with *Cla*I and *Sac*I and ligation to pCGN3212 digested with *Cla*I and *Sac*I. The resulting expression cassette pCGN3221, is digested with *Hind*III and the napin expression sequences are gel purified away and ligated to pIC20H (Marsh, *supra*) digested with *Hind*III. The final expression cassette is pCGN3223, which contains in an ampicillin resistant background, essentially identical 1.725 napin 5' and 1.265 3' regulatory sequences as found in pCGN1808. The regulatory regions are flanked with *Hind*III, *Not*I and *Kpn*I restriction sites and unique *Sal*I, *Bgl*II, *Pst*I, and *Xho*I cloning sites are located between the 5' and 3' noncoding regions. pCGN3223 is also described in WO 92/20236 which is incorporated herein by reference.

Mangosteen acyl-ACP thioesterase clone pCGN5252 (Garm FatA1) is digested with *Pst*I and *Xho*I and cloned into *Pst*I/*Xho*I digested pCGN3223 resulting in pCGN5253 containing the mangosteen thioesterase encoding region positioned for transcriptional control from the napin promoter. The *Pst*I site in pCGN5252 is located in the cloning vector at the 5' end of the cDNA clone, and the *Xho*I site is located in the 3' untranslated region at nucleotides 1233-1238.

For *Agrobacterium*-mediated plant transformation, pCGN5253 is digested with *Asp*718 and cloned into *Asp*718 digested pCGN1578, a binary vector for plant transformation, resulting in clone pCGN5255. To insert a repeat of the napin 5'/garm FatA1/napin 3' into the binary vector, pCGN5253 is digested with *Asp*718 and the napin 5'/Garm FatA1/napin 3' region gel purified. The gel purified fragment is then ligated into *Asp*718 digested pCGN5255, resulting in clone pCGN5266

pCGN5255 is transformed into *Agrobacterium tumefaciens* strain EHA101 and used to transform a high oleic acid line of *Brassica napus*. pCGN5266 is transformed into *Agrobacterium tumefaciens* strain EHA101 and used to transform *B. napus* variety Quantum (SP30021) and a low linolenic *B. napus* line, Q04 (LP004).

Example 4 Analysis of Transgenic Plants Expressing Mangosteen Thioesterase

A. Analysis of 5255-Transformants

Pooled segregating mature seeds from pCGN5255 transformants are analyzed to determine fatty acid composition by GC as described in WO 92/20236. In the 22 plants analyzed, 18:0 (stearate) levels ranged from 2.8 to 14.2 weight percent, compared to background levels in a nontransformed plant of 2.2 percent. Analyses of pooled seed samples from additional transformants were conducted, and 18:0 levels of 16.6 weight percent were detected in transformant 5255-29. Levels of 16:0 fatty acids are not affected by expression of the Garm FatA1 thioesterase.

Half seeds from transgenic plants 5255-20 and 5255-3 were similarly analyzed to determine the levels of 18:0 fatty acids obtained in individual seeds. For event 5255-20, half seeds with 18:0 fatty acid contents of up to 22.6 weight percent were obtained, and from event 5255-3, half seeds with 18:0 fatty acid contents of up to 10.0 weight percent were obtained. Similar data were obtained with other 5255 transgenic events and revealed 18:0 fatty acid levels of up to 26.6% in half seeds of 5255-29 and up to 15% in half seeds of 5255-19.

The highest half-seed plants were grown to maturity and T3 pooled seed was analyzed for fatty acid composition. Stearate levels up to 26% in pooled seed were observed in selections from transformant 5255-29. Half-seeds from the T3 pools of event 5255-20 showed individuals with stearate contents up to 39%. Selections from event 5255-3, which contains the transgene in only one locus, had T3 pooled seed with the same stearate content (11%) as the half-seeds from which they were selected, as would be expected if the insertion were homozygous. The trait is being inherited in a Mendelian fashion and the stearate levels are being maintained.

B. Analysis of 5266-Transformants

Pooled segregating mature seeds from pCGN5266 transformants are analyzed to determine fatty acid composition by GC as described in WO 92/20236. Results of these assays

for the 10 transformants in each variety with the highest stearate levels are presented in Figure 4 as weight percent fatty acid composition. In plants analyzed to date, 18:0 levels range from approximately 4 to 22 weight percent, compared to background levels in nontransformed plants of 1.4 (QO4) and 1.8 (Quantum) percent.

Half seeds from transgenic plants 5266-LP004-2 and 5266-SP30021-29 were similarly analyzed to determine the levels of 18:0 fatty acids obtained in individual seeds. Fatty acid composition data from the ten seeds of those analyzed which contained the highest levels of stearate are presented in Figure 5. For event 5266-LP004-2, half seeds with 18:0 fatty acid contents of up to 33.13 weight percent were obtained, and from event 5266-SP30021-29, half seeds with 18:0 fatty acid contents of up to 38.86 weight percent were obtained. Similar data were obtained with other 5266 transgenic events.

Example 5 Construct for Antisense Regulation of Stearoyl-ACP Desaturase in *Brassica*

Brassica napus stearoyl-ACP desaturase clones were isolated as follows. A cDNA library was constructed from RNA isolated from mid maturation (30 days after pollination) seeds of *Brassica napus* cultivar 212/86. The library was constructed using the lambda Uni-ZAP (Stratagene) vector kit according to the manufacturers directions with the following modification: 100µg of total RNA was used for cDNA synthesis, and first strand cDNA synthesis was carried out at 42°C. The cDNA library was screened with the coding region of a delta-9 desaturase gene isolated from *Brassica campestris* (Knutzon *et al.* (1992) *Proc. Natl Acad. Sci. 89*:2624-2628). Partial DNA sequence was obtained from 42 clones that hybridized with the probe. The clones fell into 5 classes. DNA sequence was obtained from the largest clone from each class. BND9 and BND16 were most closely related and 78.6% of the cDNA clones were of this class. BND11 and BND53 were related, and 19% of the cDNA clones were of this class.

An antisense gene was constructed to generate antisense RNA homologous to both of the major classes of desaturase genes, BND9 and BND11. A fragment containing the majority of

the coding region of BND9 was excised using the enzymes *Hin*dIII and *Pvu*II. The ends were filled in with the klenow fragment of DNA Polymerase I and deoxynucleotides. Plasmid pCGN3223, containing a napin expression cassette, was linearized with *Bgl*II and the ends were filled in with the klenow fragment of DNA Polymerase I and deoxynucleotides. Ligation of the pCGN3223 vector and the BND9 fragment resulted in pCGN7826. pCGN7826 was digested with *Xho*I and the ends were filled in with the klenow fragment of DNA Polymerase I and deoxynucleotides. The filled in fragment isolated from BND11 after digestion with *Bgl*II and *Pvu*II was ligated to pCGN7826 to yield pCGN7690. pCGN7690 contains a napin promoter positioned for expression of an antisense transcript homologous to both the BND9 and BND11 mRNAs. The antisense gene was excised from pCGN7690 using Asp718 and cloned into Asp718 digested binary plant transformation vector pCGN1559PASS to yield pCGN7696. pCGN1559PASS is a binary vector such as those described by McBride *et al.* (*supra*) and is prepared from pCGN1559 by substitution of the pCGN1559 linker region with a linker region containing the following restriction digestion sites: *Asp*718/*Asc*I/*Pac*I/*Xba*I/ *Bam*HI/*Swa*I/*Sse*8387(*Pst*I)/*Hin*dIII. *Asc*I, *Pac*I, *Swa*I and *Sse*8387 have 8-base restriction recognition sites and are available from New England BioLabs: *Asc*I, *Pac*I; Boehringer Manheim: *Swa*I and Takara (Japan): *Sse*8387.

Example 6 Analysis of Transgenic Plants Expressing an Antisense Desaturase Construct

Approximately 70 transgenic plants containing pCGN7696 were generated. Pools of 50 seeds were analyzed using gas chromatography to determine the fatty acid composition of the seed oils. The average level of stearate in pooled seeds from these transgenic plants was 7.7%. Pooled seeds from plant 7696-31 contained 14.7% (by weight) stearate, pooled seeds from plant 7696-54 contained 14.1% stearate, pooled seeds from plant 7696-69 contained 11.4% stearate, pooled seeds from plant 7696-36 contained 11.3% stearate, and pooled seeds from plant 7696-45 contained 9.5% stearate. By comparison, pooled seeds from untransformed control plants contain less than 3.0%

stearate. Also, in pooled seeds from 20 transgenic plants containing an antisense construct containing only the BND9 desaturase sequence under control of the napin promoter, an average of 3.6% stearate in the pooled seeds was observed. This data demonstrates the dual antisense desaturase gene construct, pCGN7696, provides for improved stearate production as compared to a single gene construct. Additional data confirm these results as the highest stearate level observed in pooled seed from a BND9 transformant was 7.6 % as compared to 14.7% with 7696 transformants. Similarly, the highest single seed level of stearate observed with BND9 transformants was 10.1% in comparison to levels of up to 29% obtained in 7696 transformants.

Segregation analysis of the oil composition of single seeds suggested that 7696-54 and 7696-31 contained two T-DNA loci, while 7696-36, 7696-69 and 7696-45 contained single T-DNA loci. Individual seeds were analyzed to determine stearate ranges for individual transformants. Stearate levels in single seeds from 7696-54 and 7696-31 ranged from 2 to 29 weight percent while single seeds from the single locus plants contained stearate contents ranging between 2 and 20 percent by weight.

Example 7 Crosses and Analysis of Plants Containing Both Antisense Desaturase and Mangosteen Thioesterase Constructs

Half-seed analysis of T2 seeds from 5255 and 7696 transformants were used to select individual transformants to be crossed for combination of the stearoyl-ACP thioesterase and antisense desaturase traits. Transgenic plants were grown from the remaining half-seed where desirable levels of 18:0 and/or numbers of GarmFatA1 gene inserts were observed. 18:0 contents of representative half seeds for 5255-3, 5255-19, 5255-29 and 5255-20 events and 7696-31, 7696-36, 7696-45 and 7696-54 events selected for crossing experiments are provided in Table 1 below.

TABLE 1

5255 Event	1/2 Seed 18:0 %	7696 Event	1/2 Seed 18:0 %
5255-29-7	21.31	7696-54-11	29.30
5255-20-11	20.33	*7696-45-1	17.65
*5255-3-29	8.11	*7696-36-1	29.62
*5255-19-1	14.95	7696-31-1	28.12

* Single insertions by segregation analysis

Data from various F1 crosses of the transgenic events indicated above are provided in Table 2. The highest stearate levels observed by half-seed analysis of seeds from selected F1 crosses are shown as weight percent of total fatty acids.

TABLE 2

Parent	5255-29-7	5255-20-11	5255-3-29	5255-19-1
7696-54-11	47.14	42.40	35.46	40.28
7696-45-1	18.77	-	7.61	13.39
7696-36-1	34.03	-	21.33	-
7696-31-1	31.84	-	-	-

The above results demonstrate the existence of numerous half seeds with stearate contents above 30% (as high as 47%) from crosses of multiple loci plants from each transgene as well as from single locus events. The combination of antisense desaturase and mangosteen TE gives higher stearate than either hemizygous parent. In some cases the F1's have higher stearate than seen so far in either homozygous parent suggesting the genes are acting synergistically.

Example 8 Transformation and Transgenic Plant Analysis with a Construct Containing Antisense Desaturase and Mangosteen Thioesterase Constructs

As an alternative to separate transformations and plant breeding as described above to obtain transgenic plants containing both antisense stearoyl-ACP desaturase and stearoyl thioesterase genes, such transgenic plants may be obtained by transformation with a single construct containing both antisense stearoyl-ACP desaturase and stearoyl thioesterase genes. One such construct is the binary transformation vector pCGN7748. pCGN7748 contains 2 copies of napin expression cassettes with antisense desaturase BND9 and BND11 genes. It also contains one copy of an expression cassette containing antisense desaturase BND9 and BND11 genes under the regulatory control of a stearoyl-ACP desaturase gene promoter, and one copy of a napin expression cassette for expression of the mangosteen thioesterase gene.

pCGN7748 was constructed by cloning two copies of the napin/antisense BND9/ antisense BND11 gene from pCGN7690 (see Example 5) as *Asp*718 fragments into *Asp*718 digested binary vector pCGN1559PASS to yield pCGN7859. The napin/mangosteen thioesterase gene was excised from pCGN5253 (see Example 3) using *Asp*718 and the ends filled in with the klenow fragment of DNA polymerase one and all 4 dNTPs. The resulting blunt ended DNA fragment was cloned into *Swa*I digested pCGN7859 to yield plasmid pCGN7743. The DNA fragment containing BND9/BND11 from pCGN7690 was excised with *Sal*I and *Xho*I, and treated with the klenow fragment of DNA polymerase one and the nucleotides dCTP and TTP and ligated to the desaturase expression cassette of pCGN5207, which had been digested with *Bam*HI and treated with klenow fragment of DNA polymerase one and the nucleotides dGTP and dATP. [pCGN5207 contains approximately 1.5Kb of the 5' regulatory region and 1.3 kB of the 3' regulatory region of the *Brassica rapa* stearoyl-ACP desaturase gene (Knutzon *et al.* (1992) *Proc. Nat. Acad. Sci. 89*:2624-2628). A polylinker containing *Bam*HI-*Pst*I-*Not*I-*Xba*I-*Nae*I-*Eco*RI-*Cla*I restriction sites separates the 5' and 3' regulatory regions in pCGN5207.] The clone resulting from

insertion of the BND9/BND11 fragment in the desaturase expression cassette was pCGN7745. The *Sma*I fragment from pCGN7745 was cloned into pCGN7743 digested with *Sse*8387 (after treating with the klenow fragment of DNA polymerase one and all 4 dNTPs) to yield pCGN7748.

pCGN7748 is transformed into *Agrobacterium tumefaciens* strain EHA101 and used to transform *B. napus* variety Quantum (SP30021).

T2 pooled seed from a plant transformed with pCGN7748 was analyzed to determine fatty acid composition and demonstrated the following fatty acid composition profile (weight percent fatty acids):

7748-SP30021-1

%16:0 = 4.8	%18:0 = 43.8	%20:0 = 6.4
%16:1 = 0.1	%18:1 = 16.6	%20:1 = 0.2
	%18:2 = 19.5	%22:0 = 1.4
	%18:3 = 7.2	

These results demonstrate that mangosteen thioesterase clone GarmFatA1 may be used to increase the 18:0 content of seed oils from transgenic plants, and that improved vegetable oils having a stearic acid content of greater than 30 weight percent may be obtained from such seeds following crushing and fatty acid extraction procedures. Furthermore, in plants expressing the mangosteen thioesterase in combination with an antisense desaturase construct, levels of 18:0 can be further increased and may exceed 50% of the total percentage of fatty acids in the plant seed oil.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

What is claimed is:

1. A method of increasing the levels of 18:0 fatty acids in plant seed triglycerides, wherein said method comprises:

growing a plant having integrated into its genome a DNA construct, said construct comprising in the 5' to 3' direction of transcription, a promoter functional in a plant seed cell, a DNA sequence encoding an acyl-ACP thioesterase protein having substantial activity on C18:0 acyl-ACP substrates and little or no activity on C16:0 acyl-ACP substrates, and a transcription termination region functional in a plant cell.

2. The method of Claim 1 wherein said plant is an oilseed crop plant.

3. The method of Claim 2 wherein said oilseed crop plant is a *Brassica* plant.

4. The method of Claim 1 wherein said thioesterase encoding sequence is from a mangosteen plant.

5. The method of Claim 4, wherein said DNA sequence encodes the mangosteen GarmFatA1 thioesterase presented in Figure 2.

6. The method of Claim 1, wherein said promoter is from a gene preferentially expressed in plant seed tissue.

7. The method of Claim 1, wherein said plant seed triglycerides comprise at least 20 mole percent C18:0 fatty acyl groups.

8. The method of Claim 1, wherein said plant seed triglycerides comprise at least 30 mole percent C18:0 fatty acyl groups.

9. The method of Claim 1, wherein said plant has integrated into its genome a second DNA construct, wherein said second DNA construct provides for antisense expression of a stearoyl-ACP desaturase gene native to said plant.

10. The method of Claim 9, wherein said plant seed triglycerides comprise at least 40 mole percent C18:0 fatty acyl groups.

11. The method of Claim 9, wherein said plant is an oilseed Brassica plant, said thioesterase encoding sequence is from a mangosteen plant and said second DNA construct is pCGN7696.

12. A plant seed comprising a minimum of 20 mole percent stearate in total fatty acids, wherein said stearate is incorporated into at least one position of a triglyceride molecule and wherein wild-type seed of said plant contains less than 5.0 mole percent stearate in fatty acids.

13. The seed of Claim 12 comprising a minimum of about 30 mole percent stearate in fatty acids.

14. The seed of Claim 12 comprising a minimum of about 40 mole percent stearate in fatty acids.

15. Plant seed oil, wherein a minimum of 20 mole percent of the acyl groups of said oil are stearate acyl groups, and wherein said oil is derived from a seed of Claim 12.

16. Plant seed oil, wherein a minimum of 40 mole percent of the acyl groups of said oil are stearate acyl groups, and wherein said oil is derived from a seed of Claim 12.

17. A *Brassica* seed comprising a minimum of 20 mole percent stearate in total fatty acids.

18. A *Brassica* seed comprising a minimum of 40 mole percent stearate in total fatty acids.

19. Plant seed oil, wherein a minimum of 20 mole percent of the acyl groups of said oil are stearate acyl groups, and wherein said oil is derived from a *Brassica* seed of Claim 17.

20. Plant seed oil, wherein a minimum of 40 mole percent of the acyl groups of said oil are stearate acyl groups, and wherein said oil is derived from a *Brassica* seed of Claim 18.

21. A construct for antisense regulation of a plant stearoyl-ACP desaturase gene, wherein said construct comprises a promoter providing for expression in a plant seed of an antisense transcript comprising DNA sequence from at least two plant stearoyl-ACP desaturase genes.

22. The construct of Claim 21, wherein said two plant stearoyl-ACP desaturase DNA sequences are from two different classes of genes from a stearoyl-ACP desaturase gene family.

23. The construct of Claim 21 represented by pCGN7696.

24. A method for increasing the stearate content in transgenic plant seeds by antisense regulation of a native stearoyl-ACP desaturase gene, the improvement of which comprises:

growing a plant comprising a construct for expression in plant seed tissues, wherein said construct provides for transcription of an antisense DNA sequence comprising portions of at least two plant stearoyl-ACP desaturase genes.

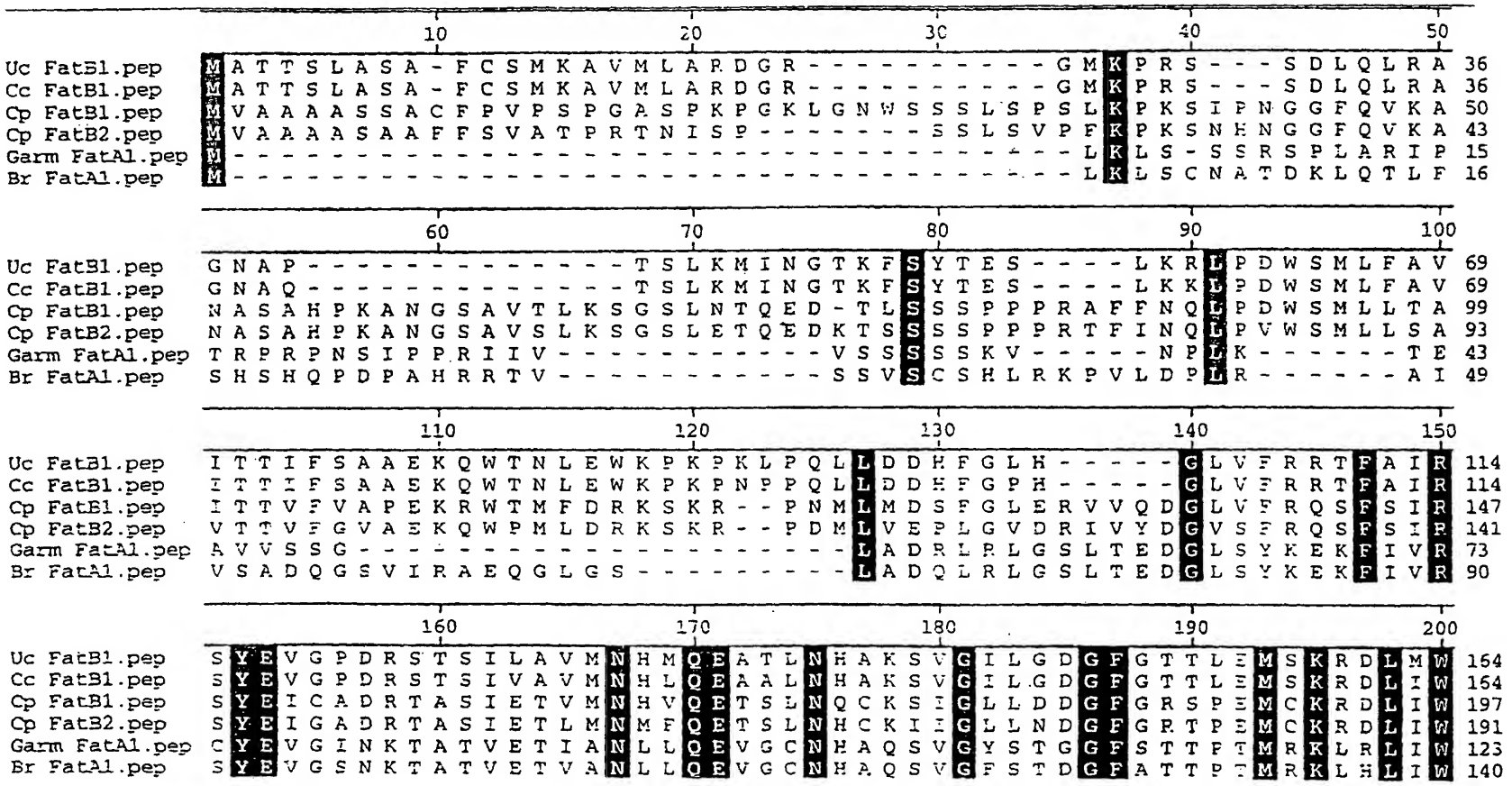
10 20 30 40 50
Uc FatB1.pep MATTSLASA-FCSMKAVMLARDGR----------GMKPRS---SDLQLRA 36
Cc FatB1.pep MATTSLASA-FCSMKAVMLARDGR----------GMKPRS---SDLQLRA 36
Cp FatB1.pep MVAAAASSACFPVPSPGASPKPGKLGNWSSSLSPSLKPKSIPNGGFQVKA 50
Cp FatB2.pep MVAAAASAAFFSVATPRTNISP--------SSLSVPFKPKSNHNGGFQVKA 43
Garm FatA1.pep M----------------------------------LKLS-SSRSPLARIP 15
Br FatA1.pep M----------------------------------LKLSCNATDKLQTLF 16
60 70 80 90 100
Uc FatB1.pep GNAP----------TSLKMINGTKFSYTES-----LKRLPDWSMLFAV 69
Cc FatB1.pep GNAQ----------TSLKMINGTKFSYTES-----LKKLPDWSMLFAV 69
Cp FatB1.pep NASAHPKANGSAVTLKSGSLNTQED-TLSSSPPPPRAFFNQLPDWSMLLTA 99
Cp FatB2.pep NASAHPKANGSAVSLKSGSLETQEDKTSSSSSPPPPRTFINQLPVWSMLLSA 93
Garm FatA1.pep TRPRPNSIPPRIIV----------VSSSSSSKV-----NPLK-------TE 43
Br FatA1.pep SHSHQPDPAHRRTV----------SSVSCSHLRKPVLDPLR------AI 49
110 120 130 140 150
Uc FatB1.pep ITTIFSAAEKQWTNLEWKPKPKLPQLLDDHFGLH-----GLVFRRTFAIR 114
Cc FatB1.pep ITTIFSAAEKQWTNLEWKPKPNPPQLLDDHFGPH-----GLVFRRTFAIR 114
Cp FatB1.pep ITTVFVAPEKRWTMFDRKSKR--PNMLMDSFGLERVVQDGLVFRQSFSIR 147
Cp FatB2.pep VTTVFGVAEKQWPMLDRKSKR--PDMLVEPLGVDRIVYDGVSFRQSFSIR 141
Garm FatA1.pep AVVSSG----------LADRLRLGSLTEDGLSYKEKFIVR 73
Br FatA1.pep VSADQGSVIRAEQGLGS----------LADQLRLGSLTEDGLSYKEKFIVR 90
160 170 180 190 200
Uc FatB1.pep SYEVGPDRSTSILAVMNHMQEATLNHAKSVGILGDGFGTTLEMSKRDLMW 164
Cc FatB1.pep SYEVGPDRSTSIVAVMNHLQEAALNHAKSVGILGDGFGTTLEMSKRDLIW 164
Cp FatB1.pep SYEICADRTASIETVMNHVQETSLNQCKSIGLLDDGFGRSPEMCKRDLIW 197
Cp FatB2.pep SYEIGADRTASIETLMNMFQETSLNHCKIIGLLNDGFGRTPEMCKRDLIW 191
Garm FatA1.pep CYEVGINKTATVETIANLLQEVGCNHAQSVGYSTGGFSTTPTMRKLRLIW 123
Br FatA1.pep SYEVGSNKTATVETVANLLQEVGCNHAQSVGFSTDGFATTPTMRKLHLIW 140

FIGURE 1A

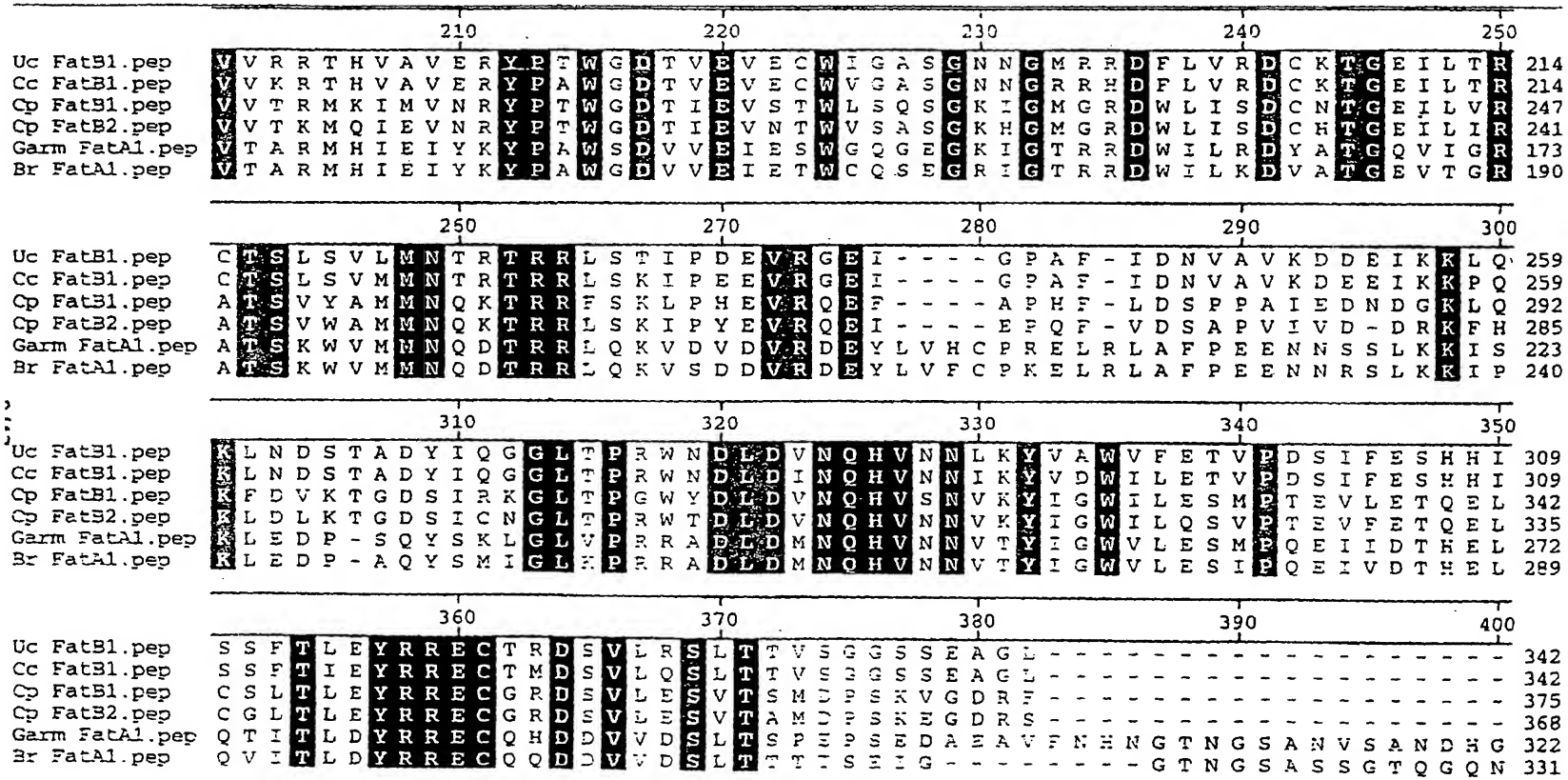
210 220 230 240 250
Uc FatB1.pep 214
Cc FatB1.pep 214
Cp FatB1.pep 247
Cp FatB2.pep 241
Garm FatA1.pep 173
Br FatA1.pep 190
250 270 280 290 300
Uc FatB1.pep 259
Cc FatB1.pep 259
Cp FatB1.pep 292
Cp FatB2.pep 285
Garm FatA1.pep 223
Br FatA1.pep 240
310 320 330 340 350
Uc FatB1.pep 309
Cc FatB1.pep 309
Cp FatB1.pep 342
Cp FatB2.pep 335
Garm FatA1.pep 272
Br FatA1.pep 289
360 370 380 390 400
Uc FatB1.pep 342
Cc FatB1.pep 342
Cp FatB1.pep 375
Cp FatB2.pep 368
Garm FatA1.pep 322
Br FatA1.pep 331

FIGURE 1B

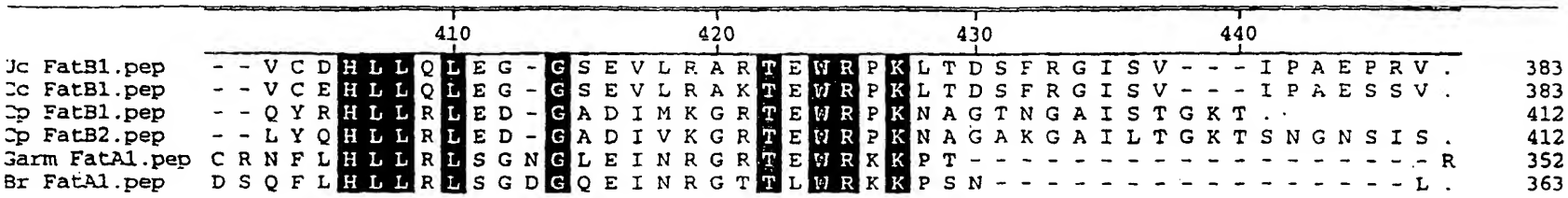
410
420
430
440
Jc FatB1.pep - - V C D H L L Q L E G - G S E V L R A R T E W R P K L T D S F R G I S V - - - I P A E P R V . 383
Cc FatB1.pep - - V C E H L L Q L E G - G S E V L R A K T E W R P K L T D S F R G I S V - - - I P A E S S V . 383
Cp FatB1.pep - - Q Y R H L L R L E D - G A D I M K G R T E W R P K N A G T N G A I S T G K T . . 412
Cp FatB2.pep - - L Y Q H L L R L E D - G A D I V K G R T E W R P K N A G A K G A I L T G K T S N G N S I S . 412
Garm FatA1.pep C R N F L H L L R L S G N G L E I N R G R T E W R K K P T - - - - - - - - - - - - - - - - - - R 352
Br FatA1.pep D S Q F L H L L R L S G D G Q E I N R G T T L W R K K P S N - - - - - - - - - - - - - - - - L . 363

FIGURE 1C

```
CCAAG ATG TTG AAG CTC TCT TCT TCC CGA AGC CCA TTG GCC CGC ATT CCC    50
      Met Leu Lys Leu Ser Ser Ser Arg Ser Pro Leu Ala Arg Ile Pro
                        5                  10                  15

ACC CGG CCC AGG CCC AAC TCC ATT CCT CCC CGG ATA ATT GTG GTT TCC    98
Thr Arg Pro Arg Pro Asn Ser Ile Pro Pro Arg Ile Ile Val Val Ser
                20                  25                  30

TCC TCA TCC AGC AAG GTT AAT CCA CTC AAA ACA GAG GCG GTG GTT TCT   146
Ser Ser Ser Ser Lys Val Asn Pro Leu Lys Thr Glu Ala Val Val Ser
            35                  40                  45

TCG GGG CTG GCT GAC CGG CTC CGG CTG GGC AGC TTG ACC GAG GAC GGG   194
Ser Gly Leu Ala Asp Arg Leu Arg Leu Gly Ser Leu Thr Glu Asp Gly
        50                  55                  60

CTT TCG TAT AAG GAG AAG TTC ATA GTG AGA TGC TAT GAG GTT GGG ATT   242
Leu Ser Tyr Lys Glu Lys Phe Ile Val Arg Cys Tyr Glu Val Gly Ile
    65                  70                  75

AAC AAG ACC GCT ACT GTT GAG ACT ATT GCC AAC CTC TTG CAG GAG GTT   290
Asn Lys Thr Ala Thr Val Glu Thr Ile Ala Asn Leu Leu Gln Glu Val
80                  85                  90
          95

GGA TGC AAT CAC GCC CAA AGC GTT GGA TAT TCG ACG GGT GGG TTT TCG   338
Gly Cys Asn His Ala Gln Ser Val Gly Tyr Ser Thr Gly Gly Phe Ser
                100                 105                 110

ACA ACC CCT ACC ATG AGA AAA TTG CGT CTG ATA TGG GTT ACT GCT CGC   386
Thr Thr Pro Thr Met Arg Lys Leu Arg Leu Ile Trp Val Thr Ala Arg
            115                 120                 125

ATG CAC ATC GAA ATC TAC AAA TAT CCA GCT TGG AGT GAT GTG GTG GAA   434
Met His Ile Glu Ile Tyr Lys Tyr Pro Ala Trp Ser Asp Val Val Glu
        130                 135                 140

ATA GAG TCG TGG GGC CAG GGT GAA GGA AAA ATC GGA ACC AGA CGT GAT   482
Ile Glu Ser Trp Gly Gln Gly Glu Gly Lys Ile Gly Thr Arg Arg Asp
    145                 150                 155

TGG ATT CTG AGA GAC TAT GCC ACT GGT CAA GTT ATT GGC CGA GCT ACT   530
Trp Ile Leu Arg Asp Tyr Ala Thr Gly Gln Val Ile Gly Arg Ala Thr
160                 165                 170                 175
```

FIGURE 2A

```
AGC AAG TGG GTA ATG ATG AAC CAA GAC ACC AGG CGA CTT CAA AAA GTC     578
Ser Lys Trp Val Met Met Asn Gln Asp Thr Arg Arg Leu Gln Lys Val
                180                 185                 190

GAT GTT GAT GTT CGT GAT GAG TAC TTG GTT CAC TGT CCA AGA GAA CTC     626
Asp Val Asp Val Arg Asp Glu Tyr Leu Val His Cys Pro Arg Glu Leu
            195                 200                 205

AGA TTG GCA TTT CCA GAG GAA AAT AAT AGC AGC TTG AAG AAA ATT TCA     674
Arg Leu Ala Phe Pro Glu Glu Asn Asn Ser Ser Leu Lys Lys Ile Ser
        210                 215                 220

AAA CTT GAA GAT CCT TCT CAA TAT TCG AAG CTG GGG CTT GTG CCT AGA     722
Lys Leu Glu Asp Pro Ser Gln Tyr Ser Lys Leu Gly Leu Val Pro Arg
    225                 230                 235

AGA GCA GAT CTG GAC ATG AAT CAA CAT GTT AAT AAT GTC ACC TAT ATT     770
Arg Ala Asp Leu Asp Met Asn Gln His Val Asn Asn Val Thr Tyr Ile
240                 245                 250                 255

GGA TGG GTG TTG GAG AGC ATG CCT CAA GAA ATC ATT GAT ACC CAT GAA     818
Gly Trp Val Leu Glu Ser Met Pro Gln Glu Ile Ile Asp Thr His Glu
                260                 265                 270

CTG CAA ACC ATA ACA TTA GAC TAC AGA CGG GAA TGC CAA CAT GAT GAT     866
Leu Gln Thr Ile Thr Leu Asp Tyr Arg Arg Glu Cys Gln His Asp Asp
            275                 280                 285

GTG GTT GAT TCC TTG ACT AGT CCA GAG CCT TCT GAA GAT GCT GAA GCA     914
Val Val Asp Ser Leu Thr Ser Pro Glu Pro Ser Glu Asp Ala Glu Ala
        290                 295                 300

GTT TTC AAC CAT AAT GGA ACA AAT GGG TCT GCA AAT GTG AGC GCC AAC     962
Val Phe Asn His Asn Gly Thr Asn Gly Ser Ala Asn Val Ser Ala Asn
    305                 310                 315

GAC CAT GGA TGC CGC AAC TTT CTG CAT CTA CTA AGA TTG TCG GGC AAT    1010
Asp His Gly Cys Arg Asn Phe Leu His Leu Leu Arg Leu Ser Gly Asn
320                 325                 330                 335

GGA CTT GAA ATC AAC CGT GGT CGT ACT GAG TGG AGA AAG AAA CCT ACA    1058
Gly Leu Glu Ile Asn Arg Gly Arg Thr Glu Trp Arg Lys Lys Pro Thr
                340                 345                 350
```

FIGURE 2B

AGA TGAGGCAATA AAGTACATTA TGTACTTTAT CGTTGCTTTA GCCGGCTTCT 1111
Arg

GGATGGTGAT TTCTTTCTGC ATTCCTTCTT TCCTTTTTGT TTTCCTAGGG TATCCTTCGC 1171

TTCTTGCCTG TAAGAGTATT ATGTTTTCCG TTTGCCCTGA AGTTGTAAAT TTGTCGAGGA 1231

ACTCGAGTCA TTGTTTGAAT CGAGGATGGT GAGAAGTGTA CTTGTTTGTT GTATTCCATT 1291

CTTCCTGAT 1300

FIGURE 2C

```
CACTCAAGAA AAAGGGCACC AATTGAACGC TACAACGGAG TAACCAAAG ATG TTT AAG    58
                                                      Met Phe Lys

ATC TCC TCT TCC CTG AGC CCA GTG GAC CAA ATC CCC CCC ATT TCC CCA    106
Ile Ser Ser Ser Leu Ser Pro Val Asp Gln Ile Pro Pro Ile Ser Pro
     5                  10                  15

CTG CCC AGG CCC AGG CCC AGG CCC ATT ACC CCC CGT GTT TTG GCC GTC    154
Leu Pro Arg Pro Arg Pro Arg Pro Ile Thr Pro Arg Val Leu Ala Val
20                  25                  30                  35

TCT TCT TCC TCT GGA AAG ATC GTT AAT AAT CCC CTT AAA GCG GAG ACT    202
Ser Ser Ser Ser Gly Lys Ile Val Asn Asn Pro Leu Lys Ala Glu Thr
                 40                  45                  50

ACG GAG GCG GTT TCC GGG GAG TTA GCG CGC CGT TTC CGG CTT GGG AGG    250
Thr Glu Ala Val Ser Gly Glu Leu Ala Arg Arg Phe Arg Leu Gly Arg
            55                  60                  65

TTG GCT GAG GAC GGG TTT TCG TAT AAG GAG AAG TTT ATA GTG AGG TGT    298
Leu Ala Glu Asp Gly Phe Ser Tyr Lys Glu Lys Phe Ile Val Arg Cys
        70                  75                  80

TAT GAG GTT GGA ATT AAC AAG ACC GCC ACT GTT GAG ACT CTT GCC AAT    346
Tyr Glu Val Gly Ile Asn Lys Thr Ala Thr Val Glu Thr Leu Ala Asn
    85                  90                  95

CTC TTA CAG GAG GTT GGA GGC AAT CAC GCC CAA AGT GTT GGA TTT TCG    394
Leu Leu Gln Glu Val Gly Gly Asn His Ala Gln Ser Val Gly Phe Ser
100                 105                 110                 115

ACG GAT GGG TTT GCG ACA ACC CAT TCC ATG AGA AAG ATG CAT CTG ATA    442
Thr Asp Gly Phe Ala Thr Thr His Ser Met Arg Lys Met His Leu Ile
                120                 125                 130

TGG GTT ACA GCT CGC ATG CAC ATA GAA ATA TAC AAA TAT CCA GCT TGG    490
Trp Val Thr Ala Arg Met His Ile Glu Ile Tyr Lys Tyr Pro Ala Trp
            135                 140                 145

AGT GAT GTG ATA GAA GTA GAG ACG TGG ATT GGG GCC GAA GGA AGA ATT    538
Ser Asp Val Ile Glu Val Glu Thr Trp Ile Gly Ala Glu Gly Arg Ile
        150                 155                 160
```

FIGURE 3A

```
GGA ACT AGA CGT AAT TGG ATT ATT AAG GAC TGT GCC ACT GAT GAA GTT      586
Gly Thr Arg Arg Asn Trp Ile Ile Lys Asp Cys Ala Thr Asp Glu Val
    165                 170                 175

ATT GGC CGA GCT ACT AGC AAG TGG GTT ATG ATG AAC CAA GAT ACC AGG      634
Ile Gly Arg Ala Thr Ser Lys Trp Val Met Met Asn Gln Asp Thr Arg
180                 185                 190                 195

CGA CTT GAA AAG GTT TCA GAT GAT GTT CGT GAG GAG CAC TTG GTT TTC      682
Arg Leu Glu Lys Val Ser Asp Asp Val Arg Glu Glu His Leu Val Phe
                200                 205                 210

AGT CCG AGA GAG CCA AGA TTG CCA TTT CCG GAT GAA AAT AAT AGC AGC      730
Ser Pro Arg Glu Pro Arg Leu Pro Phe Pro Asp Glu Asn Asn Ser Ser
            215                 220                 225

TTG AAG AAA ATT TCC AAA CTT GAC GAT CCT GCT CAA TAT TCC AAG CTA      778
Leu Lys Lys Ile Ser Lys Leu Asp Asp Pro Ala Gln Tyr Ser Lys Leu
        230                 235                 240

AGT CTT GAG CCT AGA AGA GGA GAT CTG GAC ATG AAT CAG CAT GTT AAT      826
Ser Leu Glu Pro Arg Arg Gly Asp Leu Asp Met Asn Gln His Val Asn
    245                 250                 255

AAC GTC ACC TAT ATT GGA TGG GTG TTG GAG AGC ATG CCT CAA GAA ATC      874
Asn Val Thr Tyr Ile Gly Trp Val Leu Glu Ser Met Pro Gln Glu Ile
260                 265                 270                 275

ATA GAC ACC CAT GAA CTA CAG ACA ATA ACA TTA GAC TAC CGA AGG GAA      922
Ile Asp Thr His Glu Leu Gln Thr Ile Thr Leu Asp Tyr Arg Arg Glu
                280                 285                 290

TGC CAA CAT GAT GAC TTG GTT GAT TCC TTG ACT AGT CCG GAG CCT TCT      970
Cys Gln His Asp Asp Leu Val Asp Ser Leu Thr Ser Pro Glu Pro Ser
            295                 300                 305

GAG TTC TCA GAA ACA ACA AAT GGA TCG GCA AAT GTG AGC CCC AAC GAC     1018
Glu Phe Ser Glu Thr Thr Asn Gly Ser Ala Asn Val Ser Pro Asn Asp
        310                 315                 320

AAT CGA TGC CTC AAC TTT TTG CAT CTA CTG AGA CTG TCA AGT GAT GGG     1066
Asn Arg Cys Leu Asn Phe Leu His Leu Leu Arg Leu Ser Ser Asp Gly
    325                 330                 335
```

FIGURE 3B

AGT GAA ATC AAC CGT GGT CGT ACT GTG TGG AGA AAG AAA CCT GCA AAA 1114
Ser Glu Ile Asn Arg Gly Arg Thr Val Trp Arg Lys Lys Pro Ala Lys
340 345 350 355

TGAGGCAATA ATTTACACAC TACTTAATTG TTGCTTTTTC CAGCTTCGTG TGGGTGGTGG 1174

TTTTTTTTGT TGGTTCATTT TTATGGTTTT TGGTTGGCCA TCAATTACGT TGGTGAGAAT 1234

AGTGTTATGG ATTTGGTGTG AGATTCTTTT ACATCAAAGA AACGATGTGA GATTCTTTTA 1294

CATCAAATTT TTCATAAACG 1314

FIGURE 3C

STRAIN_ID	%16:0	%18:0	%18:1	%18:2	%18:3	%20:0	%TSAT
LP004	5.23	1.84	62.91	25.15	2.04	0.80	8.33
5266-LP004-2	5.50	20.92	45.86	19.73	1.52	4.44	32.30
5266-LP004-31	5.42	17.99	51.46	17.78	1.48	3.96	28.68
5266-LP004-9	5.75	12.53	57.41	17.71	1.45	3.13	22.64
5266-LP004-7	4.56	11.50	60.36	17.26	1.36	2.96	20.34
5266-LP004-1	4.97	11.47	53.95	21.34	2.24	3.22	21.60
5266-LP004-4	5.43	10.84	55.44	21.89	1.76	2.77	20.08
5266-LP004-36	5.49	10.71	53.62	24.09	1.67	2.61	19.80
5266-LP004-20	4.41	10.26	63.98	16.00	1.18	2.37	17.95
5266-LP004-30	4.64	9.96	60.61	19.07	1.36	2.62	18.28
5266-LP004-17	4.56	9.87	63.13	16.89	1.28	2.39	17.76
SP30021	4.77	1.43	54.90	24.50	11.74	0.64	7.32
5266-SP30021-29	4.96	21.98	42.84	18.12	5.36	4.56	32.97
5266-SP30021-5	5.67	18.23	43.56	19.00	7.38	3.94	29.29
5266-SP30021-30	5.29	16.95	41.84	21.43	8.62	3.88	27.43
5266-SP30021-21	4.98	15.30	50.21	15.88	7.97	3.65	25.23
5266-SP30021-35	5.36	14.78	46.10	19.75	8.89	3.30	24.55
5266-SP30021-69	4.50	13.88	46.46	22.18	8.72	2.43	21.75
5266-SP30021-2	5.23	13.46	45.86	22.95	6.83	3.55	23.60
5266-SP30021-22	4.82	13.46	49.60	19.17	7.68	3.37	22.85
5266-SP30021-33	4.37	12.65	51.06	19.88	7.29	2.92	21.05
5266-SP30021-12	4.28	12.54	52.82	18.29	6.84	3.33	21.40

FIGURE 4

Strain ID	No	%16:0	%18:0	%18:1	%18:2	%18:3	%20:0	%Sats
5266-SP30021-29	89	3.58	38.86	36.10	10.41	2.94	6.06	50.07
5266-SP30021-29	43	4.41	33.86	35.88	14.49	3.77	5.52	45.32
5266-SP30021-29	41	4.97	33.17	36.18	14.49	3.40	5.52	45.29
5266-SP30021-29	92	4.07	32.67	27.80	21.85	5.75	5.58	44.08
5266-SP30021-29	46	4.58	31.19	33.40	17.02	4.74	6.33	44.28
5266-SP30021-29	32	4.08	30.19	40.30	14.40	3.45	5.24	41.16
5266-SP30021-29	96	4.11	29.98	32.01	18.97	5.53	6.51	42.93
5266-SP30021-29	69	3.84	29.53	37.57	16.93	3.07	6.31	41.89
5266-SP30021-29	35	3.76	29.20	36.69	18.56	2.83	6.14	41.21
5266-SP30021-29	21	3.66	29.00	48.17	8.26	2.51	5.88	40.43
5266-LP004-2	67	6.57	33.13	34.23	15.75	1.79	7.03	48.24
5266-LP004-2	6	5.79	32.46	31.28	20.68	2.19	5.89	45.80
5266-LP004-2	62	5.32	31.05	40.56	13.14	1.28	6.14	44.52
5266-LP004-2	52	5.70	29.14	37.82	18.24	1.42	5.40	41.98
5266-LP004-2	18	5.08	29.02	41.01	16.40	1.52	4.99	40.72
5266-LP004-2	56	6.45	28.48	32.60	23.09	1.84	5.47	42.12
5266-LP004-2	87	6.49	28.42	35.51	20.38	1.65	5.91	42.46
5266-LP004-2	68	6.63	27.85	31.20	24.49	1.73	5.54	41.95
5266-LP004-2	20	5.67	27.85	36.59	21.25	1.57	4.86	39.99
5266-LP004-2	96	6.17	27.77	32.75	24.49	1.82	4.82	40.33

FIGURE 5

```
AAACCTCGTA TCAAAAAAAA GAAA ATG GCA TTG AAG CTT AAC CCT TTG
                           Met Ala Leu Lys Leu Asn Pro Leu

GCA TCT CAG CCT TAC AAA CTC CCT TCC TCG GCT CGT CCG CCA ATC
Ala Ser Gln Pro Tyr Lys Leu Pro Ser Ser Ala Arg Pro Pro Ile

TCT ACT CTC AGA TCT CCC AAG TTC CTC TGC CTC GCT TCC TCT TCT
Ser Thr Leu Arg Ser Pro Lys Phe Leu Cys Leu Ala Ser Ser Ser

TCC CCT GCT CTC AGC TCC AGC ACC AAG GAG GTC GAG AGT CTG AAG
Ser Pro Ala Leu Ser Ser Ser Thr Lys Glu Val Glu Ser Leu Lys

AAG CCA TTC ACC CCA CCA AAG GAA GTC CAC GTT CAA GTC CTG CAC
Lys Pro Phe Thr Pro Pro Lys Glu Val His Val Gln Val Leu His

TCC ATG CCA CCC CAA AAG ATC GAA ATC TTC AAA TCC ATG GAA GAC
Ser Met Pro Pro Gln Lys Ile Glu Ile Phe Lys Ser Met Glu Asp

TGG GCC GAG CAC AAC CTC CTA CCT CAC CTC AAA GAC GTG GAG AAG
Trp Ala Glu His Asn Leu Leu Pro His Leu Lys Asp Val Glu Lys

TCA TGG CAG CCC CAG GAC TTC TTA CCG GAC CCT GCT TCC GAC GGG
Ser Trp Gln Pro Gln Asp Phe Leu Pro Asp Pro Ala Ser Asp Gly

TTC GAA GAC CAG GTA AAA GAG TTA AGA GAA AGA GCA AGA GAG CTC
Phe Glu Asp Gln Val Lys Glu Leu Arg Glu Arg Ala Arg Glu Leu

CCA GAT GAT TAC TTC GTT GTC TTG GTG GGT GAC ATG ATC ACA GAA
Pro Asp Asp Tyr Phe Val Val Leu Val Gly Asp Met Ile Thr Glu

GAA GCG CTT CCC ACC TAT CAA ACA ATG CTG AAC ACT TTG GAT GGT
Glu Ala Leu Pro Thr Tyr Gln Thr Met Leu Asn Thr Leu Asp Gly

GTA AGG GAT GAG ACT GGT GCT AGC CCC ACT TCA TGG GCC GTT TGG
Val Arg Asp Glu Thr Gly Ala Ser Pro Thr Ser Trp Ala Val Trp

ACT AGA GCT TGG ACT GCT GAA GAG AAT CGC CAC GGT GAT CTT TTG
Thr Arg Ala Trp Thr Ala Glu Glu Asn Arg His Gly Asp Leu Leu
```

FIGURE 6A

```
AAT AAG TAT CTT TAC TTG TCT GGT CGT GTT GAC ATG AGG CAG ATT
Asn Lys Tyr Leu Tyr Leu Ser Gly Arg Val Asp Met Arg Gln Ile

GAG AAG ACT ATT CAG TAC CTG ATT GGT TCC GGA ATG GAT CCA CGC
Glu Lys Thr Ile Gln Tyr Leu Ile Gly Ser Gly Met Asp Pro Arg

ACA GAG AAC AAC CCT TAC CTT GGC TTC ATC TAC ACC TCA TTC CAA
Thr Glu Asn Asn Pro Tyr Leu Gly Phe Ile Tyr Thr Ser Phe Gln

GAA AGA GCC ACC TTC GTC TCT CAC GGC AAC ACA GCT CGC CAA GCC
Glu Arg Ala Thr Phe Val Ser His Gly Asn Thr Ala Arg Gln Ala

AAA GAG CAC GGA GAC CTC AAG CTA GCC CAA ATC TGT GGG ACA ATA
Lys Glu His Gly Asp Leu Lys Leu Ala Gln Ile Cys Gly Thr Ile

GCT GCA GAC GAG AAG CGT CAC GAG ACG GCT TAC ACC AAG ATA GTT
Ala Ala Asp Glu Lys Arg His Glu Thr Ala Tyr Thr Lys Ile Val

GAG AAG CTT CTT GAG ATT GAT CCT GAC GGC ACT GTG GTG GCC TTT
Glu Lys Leu Leu Glu Ile Asp Pro Asp Gly Thr Val Val Ala Phe

GCG GAT ATG ATG AGG AAG AAA ATC TCG ATG CCT GCT CAC TTG ATG
Ala Asp Met Met Arg Lys Lys Ile Ser Met Pro Ala His Leu Met

TAC GAT GGG CGT GAC GAC AAG CTC TTT GAC AAC TTC TCC TCC GTG
Tyr Asp Gly Arg Asp Asp Lys Leu Phe Asp Asn Phe Ser Ser Val

GCT CAG AGG CTT GGT GTC TAC ACT GCT AAA GAT TAT GCG GAC ATT
Ala Gln Arg Leu Gly Val Tyr Thr Ala Lys Asp Tyr Ala Asp Ile

CTT GAG TTC TTG GTC GGG AGG TGG AAG ATT GAG AGC TTG AGT GGG
Leu Glu Phe Leu Val Gly Arg Trp Lys Ile Glu Ser Leu Ser Gly

CTT TCG GGT GAA GGA AAC AAA GCG CAG GAG TAT CTA TGT GGG TTG
Leu Ser Gly Glu Gly Asn Lys Ala Gln Glu Tyr Leu Cys Gly Leu

ACT CCG AGA ATC AGG AGG TTG GAT GAG AGA GCT CAA GCA AGA GCC
Thr Pro Arg Ile Arg Arg Leu Asp Glu Arg Ala Gln Ala Arg Ala
```

FIGURE 6B

AAG AAA GGA CCT AAG ATT CCT TTC AGC TGG ATA CAT GAC AGA GAA
Lys Lys Gly Pro Lys Ile Pro Phe Ser Trp Ile His Asp Arg Glu

GTG CAG CTC TGA AGAAAGGACA AAAGACATAA TAAAACCATT TTCTCTCTCT
Val Gln Leu

CTCTCCGTTC GTTATTTGAT ATGTCTGCTC TTGAAGTTGG TGTAGATTAC

TATGGTTTCT GATAATGTTC GTTGGTCTAG TTACAAAGTT GAGAAGCAGT

GTCTTAGTAA CTTTGTTTGT TTCTTTCAGT GACTTATGTT TGGTCTTTTA

GTAAACTTCT GGTAGTTAAA AACAGTTGAG CGTTTTGAGT CTGTACTCAG

TTTTCACTGT GGAGTTTTGT TCTAGTTGAA GTTAGTTTTT GTGTCAAAAA

AAAAAAAAA

FIGURE 6C

ACTCGCCGCT GTTTATTCAT AAAGCTATTC TCTATTCTGT GGACGAAACT CAACCTTTAA

AAAGGAGTCC AACCAGAGAA CGAGAGCCAG AGATAGTGTG AGAGCATTAG CCTTAGAGAG

AGAGAGAGAG AGCTTGTCTC TGAAAGAATC CACAA ATG GCA TTG AAG CTT AAC CCT
Met Ala Leu Lys Leu Asn Pro

TTG GCA TCT CAG CCT TAC AAC TTC CCT TCC TCG GCT CGT CCG CCA GTC
Leu Ala Ser Gln Pro Tyr Asn Phe Pro Ser Ser Ala Arg Pro Pro Val

TCT ACT TTC AGA TCT CCC AAG TTC CTC TGC CTC GCT TCT TCT TCT CCC
Ser Thr Phe Arg Ser Pro Lys Phe Leu Cys Leu Ala Ser Ser Ser Pro

GCT CTC AGC TCC AAG GAG GTT GAG AGT TTG AAG AAG CCA TTC ACA CCA
Ala Leu Ser Ser Lys Glu Val Glu Ser Leu Lys Lys Pro Phe Thr Pro

CCT AAG GAA GTG CAC GTT CAA GTC CTG CAT TCC ATG CCA CCC CAG AAG
Pro Lys Glu Val His Val Gln Val Leu His Ser Met Pro Pro Gln Lys

ATC GAG ATC TTC AAA TCC ATG GAA GAC TGG GCC GAG CAG AAC CTT CTA
Ile Glu Ile Phe Lys Ser Met Glu Asp Trp Ala Glu Gln Asn Leu Leu

ACT CAG CTC AAA GAC GTG GAG AAG TCG TGG CAG CCC CAG GAC TTC TTA
Thr Gln Leu Lys Asp Val Glu Lys Ser Trp Gln Pro Gln Asp Phe Leu

CCC GAC CCT GCA TCC GAT GGG TTT GAA GAT CAG GTA AGA GAG TTA AGA
Pro Asp Pro Ala Ser Asp Gly Phe Glu Asp Gln Val Arg Glu Leu Arg

GAA AGA GCA AGA GAA CTC CCT GAT GAT TAC TTC GTT GTT CTG GTG GGA
Glu Arg Ala Arg Glu Leu Pro Asp Asp Tyr Phe Val Val Leu Val Gly

GAC ATG ATC ACG GAA GAA GCG CTT CCG ACC TAT CAA ACC ATG CTG AAC
Asp Met Ile Thr Glu Glu Ala Leu Pro Thr Tyr Gln Thr Met Leu Asn

ACT TTG GAT GGA GTG AGG GAT GAA ACT GGC GCT AGC CCC ACT TCA TGG
Thr Leu Asp Gly Val Arg Asp Glu Thr Gly Ala Ser Pro Thr Ser Trp

FIGURE 7A

GCT ATT TGG ACA AGA GCT TGG ACT GCG GAA GAG AAC CGA CAC GGT GAT
Ala Ile Trp Thr Arg Ala Trp Thr Ala Glu Glu Asn Arg His Gly Asp

CTT CTC AAT AAG TAT CTT TAC TTG TCT GGA CGT GTT GAC ATG AGG CAG
Leu Leu Asn Lys Tyr Leu Tyr Leu Ser Gly Arg Val Asp Met Arg Gln

ATT GAA AAG ACC ATT CAG TAC TTG ATT GGT TCT GGA ATG GAT CCT AGA
Ile Glu Lys Thr Ile Gln Tyr Leu Ile Gly Ser Gly Met Asp Pro Arg

ACA GAG AAC AAT CCT TAC CTC GGC TTC ATC TAC ACA TCA TTC CAA GAA
Thr Glu Asn Asn Pro Tyr Leu Gly Phe Ile Tyr Thr Ser Phe Gln Glu

AGA GCC ACT TTC ATC TCT CAC GGA AAC ACA GCT CGC CAA GCC AAA GAG
Arg Ala Thr Phe Ile Ser His Gly Asn Thr Ala Arg Gln Ala Lys Glu

CAT GGA GAC CTC AAG CTA GCC CAA ATC TGC GGC ACA ATA GCT GCA GAC
His Gly Asp Leu Lys Leu Ala Gln Ile Cys Gly Thr Ile Ala Ala Asp

GAG AAG CGT CAT GAG ACA GCA TAC ACC AAG ATA GTT GAG AAG CTC TTT
Glu Lys Arg His Glu Thr Ala Tyr Thr Lys Ile Val Glu Lys Leu Phe

GAG ATT GAT CCT GAT GGT ACT GTG ATG GCG TTT GCA GAC ATG ATG AGG
Glu Ile Asp Pro Asp Gly Thr Val Met Ala Phe Ala Asp Met Met Arg

AAG AAA ATC TCG ATG CCT GCT CAC TTG ATG TAC GAT GGG CGG GAT GAA
Lys Lys Ile Ser Met Pro Ala His Leu Met Tyr Asp Gly Arg Asp Glu

AGC CTC TTT GAC AAC TTC TCT TCT GTT GCT CAG AGG CTC GGT GTT TAC
Ser Leu Phe Asp Asn Phe Ser Ser Val Ala Gln Arg Leu Gly Val Tyr

ACT GCC AAA GAC TAT GCG GAC ATT CTT GAG TTT TTG GTT GGG AGG TGG
Thr Ala Lys Asp Tyr Ala Asp Ile Leu Glu Phe Leu Val Gly Arg Trp

AAG ATT GAG AGC TTG ACC GGG CTT TCG GGT GAA GGA AAC AAA GCG CAA
Lys Ile Glu Ser Leu Thr Gly Leu Ser Gly Glu Gly Asn Lys Ala Gln

GAG TAC TTG TGT GGG TTG ACT CCG AGA ATC AGG AGG TTG GAT GAG AGA
Glu Tyr Leu Cys Gly Leu Thr Pro Arg Ile Arg Arg Leu Asp Glu Arg

FIGURE 7B

GCT CAA GCG AGA GCC AAG AAA GGA CCC AAG GTT CCT TTC AGC TGG ATA
Ala Gln Ala Arg Ala Lys Lys Gly Pro Lys Val Pro Phe Ser Trp Ile

CAC GAC AGA GAA GTG CAG CTC T AAAAAAGGAA CAAAGCTTTA ATACCTTGTC
His Asp Arg Glu Val Gln Leu

ACTCTCCATT CCCCATTTGA TCTGTCTGCT CTTGAAATTG GTGTAGATTA CTATGGTTTG

TGATAATGTT CGTGGGTCTA GTTACAAAGT TGAGAAGCAG TGATTTAGTA GCGCTTTGTT

GTTTCCAGTC TTTATATGTT TTTGTGTTTG GTCCCTTTAG TAAACTTGTT GTAGTTAAAT

CAGTTGAACT GTCTGGTCTG TACTCAGTTT TCACTGTGGA GTTTTGTTTC AGTTTGAGGT

TAGTTTCATT GCAGAGAGAA CTTCTTTATC CATTAATATG AAACTTGCTT CAAGGTAAAA

AAA

FIGURE 7C